

109652

From: Chan, Christina
Sent: Wednesday, December 03, 2003 1:57 PM
To: Davis, Minh-Tam; STIC-Biotech/ChemLib
Subject: RE: Rush search request for 09/674237

RECEIVED
 DEC -3 2003

STIC-BIOTECH/CHMLIB
 (STIC)

Pl ase rush. Thanks Chris

Chris Chan

TC 1600 New Hire Training Coordinator and SPE 1644
 308-3973
 CM-1, 9B19

-----Original Message-----

From: Davis, Minh-Tam
Sent: Wednesday, December 03, 2003 1:54 PM
To: Chan, Christina
Subject: Rush search request for 09/674237

Please search in commercial database, issued patent files, PGPUB and interference:
 SEQ ID NO:1, 2, and 3
 Thank you.
 MINH TAM DAVIS
 ART UNIT 1642, ROOM 8A01, MB 8E12
 305-2008

Searcher: _____
 Phone: _____
 Location: _____
 Date Picked Up: _____
 Date Completed: _____
 Searcher Prep/Review: _____
 Clerical: _____
 Online time: _____

TYPE OF SEARCH:
 NA Sequences: _____
 AA Sequences: _____
 Structures: _____
 Bibliographic: _____
 Litigation: _____
 Full text: _____
 Patent Family: _____
 Other: _____

VENDOR/COST (where applic.)
 STN: _____
 DIALOG: _____
 Questel/Orbit: _____
 DRLink: _____
 Lexis/Nexis: _____
 Sequence Sys.: _____
 WWW/Internet: _____
 Other (specify): _____

THIS PAGE BLANK (USPTO)



STIC SEARCH RESULTS FEEDBACK FORM

Biotech-Chem Library

Questions about the scope or the results of the search? Contact *the searcher or contact*:

Mary Hale, Information Branch Supervisor
308-4258, CM1-1E01

Voluntary Results Feedback Form

➤ I am an examiner in Workgroup: Example: 1610

➤ Relevant prior art **found**, search results used as follows:

- ☐ 102 rejection
- ☐ 103 rejection
- ☐ Cited as being of interest.
- ☐ Helped examiner better understand the invention.
- ☐ Helped examiner better understand the state of the art in their technology.

Types of relevant prior art found:

- ☐ Foreign Patent(s)
- ☐ Non-Patent Literature
(journal articles, conference proceedings, new product announcements etc.)

➤ Relevant prior art **not found**:

- ☐ Results verified the lack of relevant prior art (helped determine patentability).
- ☐ Results were not useful in determining patentability or understanding the invention.

Comments:

Drop off or send completed forms to STIC/Biotech-Chem Library CM1 - Circ. Desk



THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 4, 2003, 15:21:40 ; Search time 8605.82 Seconds
(without alignments)
1733.032 Million cell updates/sec

Title: US-09-674-237a-2
Perfect score: 3642
Sequence: 1 atgctcagttccaccaccc.....tgaccaccagcagcaatga 3642

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues
Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenBankl:*
1: gb_ba:*
2: gb_hcg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pac:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_scs:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pac:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_scs:*
28: em_un:*
29: em_vl:*
30: em_hcg_hum:*
31: em_hcg_inv:*
32: em_hcg_other:*
33: em_hcg_mus:*
34: em_hcg_pin:*
35: em_hcg_rtd:*
36: em_hcg_mam:*
37: em_hcg_vrt:*
38: em_sy:*
39: em_hngo_hum:*
40: em_hngo_mus:*
41: em_hngo_other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3640.4	100.0	3723	10	AF132478 Mus muscu
2	3639.4	99.9	5145	10	AF132481 Mus muscu
3	3229.8	88.7	4025	10	AF127798 Rattus no
4	2831	77.7	5381	9	AF114488 Homo sapi
5	2830	77.7	6439	9	AF114487 Homo sapi
6	2827.8	77.6	5287	9	AF064243 Homo sapi
7	2826.8	77.6	7247	9	AF064244 Homo sapi
8	2816.2	77.3	3812	10	AF132672 Rattus no
9	2027.6	55.7	4103	5	AF032118 Xenopus l
10	1483.8	40.7	2131	6	BD127168 Primer fo
11	1483.8	40.7	2131	9	AK074554 Homo sapi
12	1259.2	34.6	2131	6	BD158570 Homo sapi
13	1259.2	34.6	2131	9	AK027846 Homo sapi
14	1104.2	30.3	1676	6	BD127640 Primer fo
15	1104.2	30.3	1676	9	AK075290 Homo sapi
16	986.4	27.1	3241	9	HSU61166 Homo sapi
17	836.8	23.0	1299	9	BC039036 Homo sapi
18	749	20.6	1133	10	AF169621 Mus muscu
19	721	19.8	5828	6	BD167848 Method fo
20	721	19.8	5828	9	AF248540 Homo sapi
21	721	19.8	5938	9	AB033082 Homo sapi
22	713.8	19.6	4557	9	AF182199 Homo sapi
23	713.8	19.6	6091	9	AF182198 Homo sapi
24	677	18.6	4977	10	AF132479 Mus muscu
25	673.8	18.5	3594	10	AF132480 Mus muscu
26	648	17.8	4447	6	AX428899 Sequence
27	617.4	17.0	5804	10	AK122480 Mus muscu
28	531.4	14.6	78190	2	AC106760 Homo sapi
29	531.4	14.6	179264	2	AC023112 Homo sapi
30	513.8	14.1	677	6	AX677983 Sequence
31	485.4	13.3	877	6	BD124639 Primer fo
32	485.4	13.3	877	6	BD126208 Primer fo
33	477.4	13.1	2017	6	AR238090 Sequence
34	477.4	13.1	2017	6	AR257631 Sequence
35	477.4	13.1	2017	6	AR283677 Sequence
36	477.4	13.1	2017	6	AX36305 Sequence
37	435	11.9	475	10	AY127576 Mus muscu
38	387.4	10.6	1329	6	BD157252 Primer fo
39	387.4	10.6	1329	9	AK021545 Homo sapi
40	372	10.2	493	10	AF525079 Mus muscu
41	360.8	9.9	2085	9	BC038963 Homo sapi
42	342.2	9.4	1996	9	AF180522 Homo sapi
43	332.4	9.1	2873	6	AR175271 Sequence
44	331.4	9.1	1391	9	AF001630 Homo sapi
45	331.4	9.1	2004	9	AF038189 Homo sapi

ALIGNMENTS

RESULT 1
AF132478
LOCUS AF132478 3723 bp mRNA linear ROD 09-MAR-1999
DEFINITION Mus musculus Ebel protein mRNA, complete cds.
ACCESSION AF132478
VERSION AF132478.1 GI:4378884
KEYWORDS
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
Sengar, A.S., Wang, W., Bishop, V., Cohen, S. and Egan, S.B.
The EH and SH3 domain Eee proteins regulate endocytosis by linking
TITLE to dynamin and Eps15

JOURNAL: EMBO J. 18 (5), 1159-1171 (1999)
 MEDLINE: 99164083
 PUBMED: 10064583
 REFERENCE: 2 (bases 1 to 3723)
 AUTHORS: Bengt, A.S., Wang, W., Cohen, S., Bishop, J. and Egan, S.E.
 TITLE: Direct Submission
 JOURNAL: Submitted (02-MAR-1999) Programs in Cancer & Blood
 Research/Developmental Biology, The Hospital for Sick Children,
 University Avenue, Toronto, ON M5G-1X8, Canada

FEATURES

SOURCE

CDS

```

BASE COUNT      1065 a      940 c      1026 g      692 t
ORIGIN
/codon_start=1
/product="Ese1 protein"
/protein_id="ABD19746.1"
/db_xref="GI:4378885"
/translation="MAQFPYPRGSLDWYALTVEERAHDOQFLSLKPIAGITLTPV
KNFFPQSLQPVLAQIWTALDMMNDGMDQVERISIAKLILKLIQGIQVLESTLPPV
KQOPVALLSAPAFGIGIASMPPLTAAVPMQSPVPMGSLPVVSGPLVASVPLPLA
GAPVITQPLPAFAPHAATWMTGSSPSRSGSPQNLTKQKQSPVNASAPDAEALVY
OSSRLKYQLPNSHDKTMSGLTSGTQATLTIMQSLIPQAOLASINLSDIAQDLATV
EEFLIAMHLIDVANGGQPLPVLPPEYIPSEFRALISGMSVSSSDVDRLPPEEP
SEDDQOPEKTKPYTPEDKKRPNFERGSGVLEKKRRALLQORKEERAPLQAPQEE
KEREROEBAKQOLELEKQLEKQRELEKQREBERKELERREBAKRELEORQLEME
NRROSLNQRNKNEDEGIVLQARRKTLLELBALDKKHQLEGLQDILRCALTORQI
IESTNKSESLIAELTTHLQQLQESQWMLGRLIPKQLISLQKQVQDNLHRSLSL
LKRALIEAKELAQQLREQLDEVERETRSKQEIIVFNQQLLELRISHKQQLQORSI
BAARKIQKEORRSKLELEKQEKQQRVROEKDQWLEHVQOEBQRPKPHSEBRLLK
EDSVRKAEEERAKEMQDKOSRLPHPOEOPAKTATAPWSTTEKGPLTISQESVYK
VYVALYVPEERSHDELITOPGDIYVMDSESOTGEBGWGLKKGKTMFPANVAYEKI
ENEVPYPAKPYTDLTSLAPAKLALRETAAPLPYTSSESTTPNNMADDSSTWPSNN
KPEYDMDWTMAQOBSLTVPSAQQLKQVQKQLESPSLVQLQSGKVASLQAQAL
YPMRAKKNNHLNPNKSDIVLYALEQQDMWPFEBVQOQMKPFKSYKLISGVKRSIT
DTGTPSEPSALKEVRVSPAPKAPLAPGEEELIAMYTYSSEQDGLTFOGQVAVI
WMTGTVGDSGVFPENYVRLKDSBEGSGAGTYSIGKKPELIAOVASVAAAGPEVLA
APGQILIRKKNPQGMWEGELQARCKGQITMFPANVYKLISPGSKITPHELPTAA
AQACQVQYCTMDYTAQNDELAFASKQILINLANKEDPMWKGVSQVGLFPPSNYKL
TTDMDPSQO"

```

Query Match	100.0%	Score 3640.4	DB 10	Length 3723
Best Local Similarity	100.0%	Pred. No. 0		
Matches 3641	Conservative	0	Mismatches 1	Indels 0
			Gaps	0
QY	1	ATGGCTCAGTTTCCACACCTTTCGGTGGTAGCTCGATGTCTGGGCCATPACTGTGGAG	60	
Db	1	ATGGCTCAGTTTCCACACCTTTCGGTGGTAGCTCGATGTCTGGGCCATPACTGTGGAG	60	
QY	61	GAAAGGGCCAAAGCATGACCAAGCATTTCTTTAGCTCGAAGCCGATATGCGGATTTATTTACT	120	
Db	61	GAAAGGGCCAAAGCATGACCAAGCATTTCTTTAGCTCGAAGCCGATATGCGGATTTATTTACT	120	
QY	121	GATGATCAAGCGAGGAACTTTTTTTTCCAAATCTGGGTTACTCGACCTGTCTTAGACAATA	180	
Db	121	GATGATCAAGCGAGGAACTTTTTTTTCCAAATCTGGGTTACTCGACCTGTCTTAGACAATA	180	
QY	181	ATATGGGCGCTAGCGGACATGAATAACGATGGAGAGATGATCAAGTGAATTTTTCATA	240	
Db	181	ATATGGGCGCTAGCGGACATGAATAACGATGGAGAGATGATCAAGTGAATTTTTCATA	240	
QY	241	GCCATGAAGCTTATCAAACTGAAGCTACAAGGATATCACTCCCTCCACACTTTCGCCCT	300	
Db	241	GCCATGAAGCTTATCAAACTGAAGCTACAAGGATATCACTCCCTCCACACTTTCGCCCT	300	
QY	301	GTCATGAACAACAGCAACCAAGTGGCTATTTTCCAGTGCACCGACTTTTGGTATAGAGGAGATT	360	
Db	301	GTCATGAACAACAGCAACCAAGTGGCTATTTTCCAGTGCACCGACTTTTGGTATAGAGGAGATT	360	
QY	361	GCTAGCATGCCACCACTCAAGCTGTGTGCTCTGTGCCAATGGGCTCAATTCAGAGTTT	420	

Db	361	GCATGCAATGCCACACATCAAGCTGTGGTCTCTGTGCAATGGGCTCATTCCAGTTGTT	420
Qy	421	GGAAATGTCCTCACCCTTAGTATCTTCTGTCCCTTCAGACAGAGTCCCTCCCTGGCTAAC	480
Db	421	GGAAATGTCCTCACCCTTAGTATCTTCTGTCCCTTCAGACAGAGTCCCTCCCTGGCTAAC	480
Qy	481	GGGGCTCTCCCGGTCATACAGCCCTTCGCGCTTGGTGGCATCCTGTAGGCACATGGGCA	540
Db	481	GGGGCTCTCCCGGTCATACAGCCCTTCGCGCTTGGTGGCATCCTGTAGGCACATGGGCA	540
Qy	541	AAGAGTCTTCTCCTCAGACAGATCGTGTCAGAGGTCACAAATTAAACATTAATTACAGAG	600
Db	541	AAGAGTCTTCTCCTCAGACAGATCGTGTCAGAGGTCACAAATTAAACATTAATTACAGAG	600
Qy	601	GCACAAATCATTTCCATGTGCGCCAGCGCCCTTCAGACAGCAAGATGGGCTGTGCTCACTCA	660
Db	601	GCACAAATCATTTCCATGTGCGCCAGCGCCCTTCAGACAGCAAGATGGGCTGTGCTCACTCA	660
Qy	661	TCAAGGCTGAAATTCAGGCACTTATTCACAGCCACGACAAAATATAGTGGCACTTAA	720
Db	661	TCAAGGCTGAAATTCAGGCACTTATTCACAGCCACGACAAAATATAGTGGCACTTAA	720
Qy	721	ACAGGTCGCCGAGCAAACTATTCTCATGCAATCAAGTTTAACTCCAGGCTCAGCTGGCT	780
Db	721	ACAGGTCGCCGAGCAAACTATTCTCATGCAATCAAGTTTAACTCCAGGCTCAGCTGGCT	780
Qy	781	TCAATATGGAATCTTTCTGACATTGATCAAGATGGAATACTCACTGCAGAGAAATTTATC	840
Db	781	TCAATATGGAATCTTTCTGACATTGATCAAGATGGAATACTCACTGCAGAGAAATTTATC	840
Qy	841	CTTACTATGCACTTAATTGATGTGGCCATGTCTGTGTCAGACATCGCGCCCGCTCTGCT	900
Db	841	CTTACTATGCACTTAATTGATGTGGCCATGTCTGTGTCAGACATCGCGCCCGCTCTGCT	900
Qy	901	CCAGAAATACATCCCTCCCTCCTTCCAGAAAGATTCGCTCCGGCAGTGGGATGTCGCTATA	960
Db	901	CCAGAAATACATCCCTCCCTCCTTCCAGAAAGATTCGCTCCGGCAGTGGGATGTCGCTATA	960
Qy	961	AGCTCTTCTTCTGTGATCAAGAGGCTGTCAGAGGCGCTGTCTCAGAGATGAGCAGCAG	1020
Db	961	AGCTCTTCTTCTGTGATCAAGAGGCTGTCAGAGGCGCTGTCTCAGAGATGAGCAGCAG	1020
Qy	1021	CCAGAGAAAGAACTGCTCTGTGACATTTGAGATTAAGAACGGGAGAACTTTCAGCGAGGC	1080
Db	1021	CCAGAGAAAGAACTGCTCTGTGACATTTGAGATTAAGAACGGGAGAACTTTCAGCGAGGC	1080
Qy	1081	AGTGTGAGGCTGAGGAAGCGCGCCAGAGCGCTCTTGGAGCAGCGCCAAAGACAGAG	1140
Db	1081	AGTGTGAGGCTGAGGAAGCGCGCCAGAGCGCTCTTGGAGCAGCGCCAAAGACAGAG	1140
Qy	1141	CGGTTGGCTCAGCTGAGCGCGCCAGACAGAGAGGAAGACCGGAGCGCCAGAGCAG	1200
Db	1141	CGGTTGGCTCAGCTGAGCGCGCCAGACAGAGAGGAAGACCGGAGCGCCAGAGCAG	1200
Qy	1201	GAGGCCAAGCGGCACTGTGAGCTGAGAGAGCAGCTTGAAGACGACGGGAGCTTGAAGCGG	1260
Db	1201	GAGGCCAAGCGGCACTGTGAGCTGAGAGAGCAGCTTGAAGACGACGGGAGCTTGAAGCGG	1260
Qy	1261	CAGGAGAGGAGGAGAGGAGGAAGAGATGAGAGGGGCGGAGCGCCGAAAAAGGGAAC	1320
Db	1261	CAGGAGAGGAGGAGAGGAGGAAGAGATGAGAGGGGCGGAGCGCCGAAAAAGGGAAC	1320
Qy	1321	GAAAGGCAAGCAAACTTTGAATGGGAAACGGAACCGGAGACAGGAATCTTGAATCAGAGG	1380
Db	1321	GAAAGGCAAGCAAACTTTGAATGGGAAACGGAACCGGAGACAGGAATCTTGAATCAGAGG	1380
Qy	1381	AACAAGAGCAGAGAGGCAACCGTGTCTTGAAGCAAGAGAGGAAGACTTGAAGTTTGAAG	1440
Db	1381	AACAAGAGCAGAGAGGCAACCGTGTCTTGAAGCAAGAGAGGAAGACTTGAAGTTTGAAG	1440
Qy	1441	TTAAGAGCTTGAAATGACAAAAGCATCAGTTAAGAGAAAATTTCAAGATATCAGGTGT	1500

Db	1441	TTTAGAGCTCTGAAATGACAAAAGATCAGCTTAGAAGGAAAATTCCAGAGATATCAGGTGT	15000
OY	1501	CGACTGGCAACCCAGAGGCGAAGAAATTGAGAGCAGAA CAAGTCTAGAGCTTAAGAAATT	15600
Db	1501	CGACTGGCAACCCAGAGGCGAAGAAATTGAGAGCAGAA CAAGTCTAGAGCTTAAGAAATT	15600
OY	1561	GCTGAATACCCACTTACAGAGAGATTTGCAGGAATCTCAGCAAAATCTTGGAAAGCTT	16200
Db	1561	GCTGAATACCCACTTACAGAGAGATTTGCAGGAATCTCAGCAAAATCTTGGAAAGCTT	16200
OY	1621	ATTCCAGAGAAACAGATACTCAGTACAGATTAAACAAGTCCAGCAGAA CAGTTTGCAAT	16800
Db	1621	ATTCCAGAGAAACAGATACTCAGTACAGATTAAACAAGTCCAGCAGAA CAGTTTGCAAT	16800
OY	1661	AGAGACTGCTTTTACCCTCAAAAAGCCTTGGAGCAAAAGAGCTGGCCCGCAGCAG	17400
Db	1661	AGAGACTGCTTTTACCCTCAAAAAGCCTTGGAGCAAAAGAGCTGGCCCGCAGCAG	17400
OY	1741	CTCCGGGAGCAGCTGGAGAGGTGGAGAGAGACCAAGTCAAAAGCTGCAGAGATTGAT	18000
Db	1741	CTCCGGGAGCAGCTGGAGAGGTGGAGAGAGACCAAGTCAAAAGCTGCAGAGATTGAT	18000
OY	1801	GTTTTCAACAAC CAGCTGAAGGAACTGAGAGATACATAGCAAA CACAATCTCCAGAG	18600
Db	1801	GTTTTCAACAAC CAGCTGAAGGAACTGAGAGATACATAGCAAA CACAATCTCCAGAG	18600
OY	1861	CAGAGGTCCTTGAGGCAAGCCGACCTGAAGCAGAAAGAGCAGAGAGAGAGCTTGAG	19200
Db	1861	CAGAGGTCCTTGAGGCAAGCCGACCTGAAGCAGAAAGAGCAGAGAGAGAGCTTGAG	19200
OY	1921	TTTAGAAGAGCAAAAGAGAGAGCTTCAGAGACGATTCAAGAAAGGGA CAGCAATGCTGT	19800
Db	1921	TTTAGAAGAGCAAAAGAGAGAGCTTCAGAGACGATTCAAGAAAGGGA CAGCAATGCTGT	19800
OY	1981	GAGCATGTGCAAGAGAGAGAGCAGCCCGGAGAAACCCCAAGAGAGACAGACTG	20400
Db	1981	GAGCATGTGCAAGAGAGAGAGCAGCCCGGAGAAACCCCAAGAGAGAGACTG	20400
OY	2041	AAGAGGGAGAGCAGTGTGACGAGAAAGAGGCGAAGAGAGAGCCAAACCGGAAATGCA	21000
Db	2041	AAGAGGGAGAGCAGTGTGACGAGAAAGAGGCGAAGAGAGAGCCAAACCGGAAATGCA	21000
OY	2101	GACAAAGCAGAGTCGGCTTTTCCATTCGATCAGAGCCAGCTAAGCTGGCCACCAGGCA	21600
Db	2101	GACAAAGCAGAGTCGGCTTTTCCATTCGATCAGAGCCAGCTAAGCTGGCCACCAGGCA	21600
OY	2161	CCCTGGTTCACACAGAGAAAGCCCGCTTACATTCGACACAGAGAGTGTAAAGTG	22200
Db	2161	CCCTGGTTCACACAGAGAAAGCCCGCTTACATTCGACACAGAGAGTGTAAAGTG	22200
OY	2221	GTAATTTACGAGGCGCTGTACCCCTTTGAAATCCAGAGTCAAGATGATCACCATCCAG	22800
Db	2221	GTAATTTACGAGGCGCTGTACCCCTTTGAAATCCAGAGTCAAGATGATCACCATCCAG	22800
OY	2281	CCAGAGATATATAGTCATGATGATGAAAGCCAGACTGAGAGCCAGATGGCTTGGAGGA	23400
Db	2281	CCAGAGATATATAGTCATGATGATGAAAGCCAGACTGAGAGCCAGATGGCTTGGAGGA	23400
OY	2341	GAGCTGAAAGGGAGAACGGAGATGGTTCCTTGCAAACTATGCAAGAAATTCAGAAAT	24000
Db	2341	GAGCTGAAAGGGAGAACGGAGATGGTTCCTTGCAAACTATGCAAGAAATTCAGAAAT	24000
OY	2401	GAGGTTCCCACTCCAGCCAAAC CAGTGAACGATCTGAATCTGCCCCCTGCCCCCAATCTG	24600
Db	2401	GAGGTTCCCACTCCAGCCAAAC CAGTGAACGATCTGAATCTGCCCCCTGCCCCCAATCTG	24600
OY	2461	GCTCTGGGTGAGACCCCTGCTCTTGGCACTGACCTTTCTGAGCCCTCCACAACCCCC	25200
Db	2461	GCTCTGGGTGAGACCCCTGCTCTTGGCACTGACCTTTCTGAGCCCTCCACAACCCCC	25200
OY	2521	AAACAATGGGCGAGATTTCAATTTCCACGTGGCCCAAGCTCAAAAGAGAGCCAGAAACG	25800
Db	2521	AAACAATGGGCGAGATTTCAATTTCCACGTGGCCCAAGCTCAAAAGAGAGCCAGAAACG	25800

QY	2581	GACAACTGGGATACGTGGGGGGGCTCAGCTTCTCTGACCCGATACCTAAGCTGGGACAGTTA	2643
Db	2581	GACAACTGGGATACGTGGGGGGGCTCAGCTTCTCTGACCCGATACCTAAGCTGGGACAGTTA	2643
QY	2641	CGGACAGATCAGGCTTTTACCCGACGACAGCCACTGGCTCTCCGACTCTCCGCTCGT	2700
Db	2641	CGGACAGATCAGGCTTTTACCCGACGACAGCCACTGGCTCTCTCCGACTCTCCGCTCGT	2700
QY	2701	GGCCAGGGGTGA AAAAGGTGGAAAGGGCTTCAAGCCGAACCCCTGTATTCCTTGGAGGCCAA	2760
Db	2701	GGCCAGGGGTGA AAAAGGTGGAAAGGGCTTCAAGCCGAACCCCTGTATTCCTTGGAGGCCAA	2760
QY	2761	AAAGACAACCACTTAAATTTTAA CA AAAAGTGAAGCTCATCAACCGTTCGTGAAACAGCAAGAC	2820
Db	2761	AAAGACAACCACTTAAATTTTAA CA AAAAGTGAAGCTCATCAACCGTTCGTGAAACAGCAAGAC	2820
QY	2821	ATGTGTGCTTTGGAGAAAGTTCAAGGTCAAGAGGGTTGGTTCCCAAGTCTTACGTGAAA	2880
Db	2821	ATGTGTGCTTTGGAGAAAGTTCAAGGTCAAGAGGGTTGGTTCCCAAGTCTTACGTGAAA	2880
QY	2881	CTCAATTCAGGGGCCGATAGGAATTCACAAGATCGATCTGGCCCTACTGAAAGTCT	2940
Db	2881	CTCAATTCAGGGGCCGATAGGAATTCACAAGATCGATCTGGCCCTACTGAAAGTCT	2940
QY	2941	GCTAGTCTMAAGAGAGTGGCTTCCCGCGCGCCAGGCAGCATTTCCCGAGAAAGATTT	3000
Db	2941	GCTAGTCTMAAGAGAGTGGCTTCCCGCGCGCCAGGCAGCATTTCCCGAGAAAGATTT	3000
QY	3001	ATTGCCATGTACACATACAGAGATTTCTGAGCAGAGAAATTTAACTTTCAGCAAGGGAT	3060
Db	3001	ATTGCCATGTACACATACAGAGATTTCTGAGCAGAGAAATTTAACTTTCAGCAAGGGAT	3060
QY	3061	GTGATTTGGTTCACCAAGAAAGATGTGATCTGTGTGACCGGAA CGGTGGCGCA CAAGTCC	3120
Db	3061	GTGATTTGGTTCACCAAGAAAGATGTGATCTGTGTGACCGGAA CGGTGGCGCA CAAGTCC	3120
QY	3121	GGAGTCTTCCCTTCTTAACTATGTAGGCTTAAAGATTCAGAGGGCTCTGAACTGTGGG	3180
Db	3121	GGAGTCTTCCCTTCTTAACTATGTAGGCTTAAAGATTCAGAGGGCTCTGAACTGTGGG	3180
QY	3181	AAAACAGGGAGTTTAGAAAAAAAACCTGAAATTTGCCAGGTTATTTGCTTCTAGCTGCT	3240
Db	3181	AAAACAGGGAGTTTAGAAAAAAAACCTGAAATTTGCCAGGTTATTTGCTTCTAGCTGCT	3240
QY	3241	ACTGGTCCCGAACAACCTACCCCTGGCTCTTGGGAGCTGATTTCTGATCCGAAAAAAGAC	3300
Db	3241	ACTGGTCCCGAACAACCTACCCCTGGCTCTTGGGAGCTGATTTCTGATCCGAAAAAAGAC	3300
QY	3301	CCAGGTGATGTGTGGAAAGAGAACTGCAGACTTCGAGGAAAAAAGCGCAGATTTGGGTGG	3360
Db	3301	CCAGGTGATGTGTGGAAAGAGAACTGCAGACTTCGAGGAAAAAAGCGCAGATTTGGGTGG	3360
QY	3361	TTTCCAGCAATTTATGTCAAACTTCTPAAGCCCGGAAACAAAGCAAAATTCACCCCAACTGAG	3420
Db	3361	TTTCCAGCAATTTATGTCAAACTTCTPAAGCCCGGAAACAAAGCAAAATTCACCCCAACTGAG	3420
QY	3421	CTACCCAAAGACCGCAGGTGCAGCCAGCTGTCAGGTGATCCGAGTGTACGATTTACACC	3480
Db	3421	CTACCCAAAGACCGCAGGTGCAGCCAGCTGTCAGGTGATCCGAGTGTACGATTTACACC	3480
QY	3481	GCCCAAGAACGATGACGAACTAAGCCTTACGAAAGGCGCAGATCATCAACGTCCTAACAG	3540
Db	3481	GCCCAAGAACGATGACGAACTAAGCCTTACGAAAGGCGCAGATCATCAACGTCCTAACAG	3540
QY	3541	GAGGACCCGGAAGTGTGGAAAGGAGAAAGTCAAGTGGGCAAGTTGGGCTCTTCCATTCAT	3600
Db	3541	GAGGACCCGGAAGTGTGGAAAGGAGAAAGTCAAGTGGGCAAGTTGGGCTCTTCCATTCAT	3600
QY	3601	TATGTAAAGCTGACCAACAGACATGAGACCCCAAGCCAGCAATGA	3642
Db	3601	TATGTAAAGCTGACCAACAGACATGAGACCCCAAGCCAGCAATGA	3642

Dp	1081	AGTGTGAGCTGAGAAAGCCCGGACAGCGCTCTTTGAGCGACGACCGCAAAAGACGAGG	1140
Qy	1141	CGGTGTGGCTCAGCTGAGCGCGCCGACAGAGAGAGAAAGACGGGAGCGCCAGAGCAG	1200
Dp	1141	CGGTGTGGCTCAGCTGAGCGCGCCGACAGAGAGAGAAAGACGGGAGCGCCAGAGCAG	1200
Qy	1201	GAGGCTAAAGCGGACGCTGAGCTGGAGAAAGCAGCTGGAGAACGACGGGAGCTGAGCGG	1260
Dp	1201	GAGGCTAAAGCGGACGCTGAGCTGGAGAAAGCAGCTGGAGAACGACGGGAGCTGAGCGG	1260
Qy	1261	CAGCGAGAGGAGAGAGAGAGAGATTCAGAGAGCGCGAGCGCCGCAAAACGGGAACCTG	1320
Dp	1261	CAGCGAGAGAGAGAGAGAGAGATTCAGAGAGCGCGAGCGCCGCAAAACGGGAACCTG	1320
Qy	1321	GAAAGGCGACGACCACTTGAATGGGAAACGGAACCGGAGACAGAACTCTGAATCAGAGG	1380
Dp	1321	GAAAGGCGACGACCACTTGAATGGGAAACGGAACCGGAGACAGAACTCTGAATCAGAGG	1380
Qy	1381	AACAAAGAGCAGAGAGGGGCAACCGTGTCTTGAAGCGAAGAGAAAGACTCTGAGCTTGAG	1440
Dp	1381	AACAAAGAGCAGAGAGGGGCAACCGTGTCTTGAAGCGAAGAGAAAGACTCTGAGCTTGAG	1440
Qy	1441	TTAGAAGGCTCGAATGACAAAANAAGATAGCTAGAAGGAAAACCTTCAGAGATACAGGT	1500
Dp	1441	TTAGAAGGCTCGAATGACAAAANAAGATAGCTAGAAGGAAAACCTTCAGAGATACAGGT	1500
Qy	1501	CGACTGCGAACCAGAGGCGAAGAAATTGAGAGCAGAACAGTCTAGAGACTTAAGATT	1560
Dp	1501	CGACTGCGAACCAGAGGCGAAGAAATTGAGAGCAGAACAGTCTAGAGACTTAAGATT	1560
Qy	1561	GCTGAATCACCACCTTACAGCAGCAGTTGAGAGAAATCAGCAATGCTTGAAGACTT	1620
Dp	1561	GCTGAATCACCACCTTACAGCAGCAGTTGAGAGAAATCAGCAATGCTTGAAGACTT	1620
Qy	1621	ATTCAGAGAAACAGATACAGTACGAGCCAGTTAAACAAGTCCAGCAGAAACAGTTTGAT	1680
Dp	1621	ATTCAGAGAAACAGATACAGTACGAGCCAGTTAAACAAGTCCAGCAGAAACAGTTTGAT	1680
Qy	1681	AGAGACTGCTTCTTACCTCCTCAAAAAGACCTTGGAGCAAGAGAGACTGGCCCGGACAG	1740
Dp	1681	AGAGACTGCTTCTTACCTCCTCAAAAAGACCTTGGAGCAAGAGAGACTGGCCCGGACAG	1740
Qy	1741	CTCCGGGAGCAGCTGAGCAGAGGTGAGAGAGACCAGGTCAAAAGCTGACAGAGATTGAT	1800
Dp	1741	CTCCGGGAGCAGCTGAGCAGAGGTGAGAGAGACCAGGTCAAAAGCTGACAGAGATTGAT	1800
Qy	1801	GTTTTCACCAACACAGCTGAGAGAACTGAGAGATACCTAGCAAAACAGCACTCCAGAG	1860
Dp	1801	GTTTTCACCAACACAGCTGAGAGAACTGAGAGATACCTAGCAAAACAGCACTCCAGAG	1860
Qy	1861	CAGAGGTCCCTGAGAGGCGAGCGGACTGAAGGAGAAAGCGAGAGAGGAGAGCTGAG	1920
Dp	1861	CAGAGGTCCCTGAGAGGCGAGCGGACTGAAGGAGAAAGCGAGAGAGGAGAGCTGAG	1920
Qy	1921	TTAGAGAAAGCAAAAGAGAGAGCTCAGAGAGAGTTTCAGAAAGGACCAAGCAATGCTG	1980
Dp	1921	TTAGAGAGCAAAAGAGAGAGAGCTCAGAGAGAGAGTTTCAGAAAGGAGCAAGCAATGCTG	1980
Qy	1981	GAGCATGTGACAGAGAGAGACAGCACGCCCCGGAACCCACAGAGAGACAGACTG	2040
Dp	1981	GAGCATGTGACAGAGAGAGACAGCACGCCCCGGAACCCACAGAGAGACAGACTG	2040
Qy	2041	AAGAGGGAGAGACAGTGTACAGAGAAAGAGGCGGAAAGAGAGGCAAGCGGGAATCCAA	2100
Dp	2041	AAGAGGGAGAGACAGTGTACAGAGAAAGAGGCGGAAAGAGAGGCAAGCGGGAATCCAA	2100
Qy	2101	GACAAAGCAGAGTCCGCTTTTCCATCCGAGTACAGAGCCAGCTAAGCTTGCCACCCAGGCA	2160
Dp	2101	GACAAAGCAGAGTCCGCTTTTCCATCCGAGTACAGAGCCAGCTAAGCTTGCCACCCAGGCA	2160
Qy	2161	CCCTGTGTACACAGAGAAAGGCCCGCTTACATTTCTGACAGAGAGTGTAAAGTG	2220
Dp	2161	CCCTGTGTACACAGAGAAAGGCCCGCTTACATTTCTGACAGAGAGTGTAAAGTG	2220

QY	2221	GTATATTACCGAGGGCTGTACCCCTTTGGAATCCGAAGTCAAGATGATACATCCAG	2280
Db	2221	GTATATTACCGAGGGCTGTACCCCTTTGGAATCCGAAGTCAAGATGATACATCCAG	2280
QY	2281	CCAGAGATATTAGTCATGATGGATGGAAGGCCAGACTGGAGAGCCAGGATGGCTTGGAGGA	2340
Db	2281	CCAGAGATATTAGTCATGATGGATGGAAGGCCAGACTGGAGAGCCAGGATGGCTTGGAGGA	2340
QY	2341	GAGCTGAAGGGAAGACGGGATGGTTCCTTGCAAACTATGCAAAAAGATTCCAGAAAAT	2400
Db	2341	GAGCTGAAGGGAAGACGGGATGGTTCCTTGCAAACTATGCAAAAAGATTCCAGAAAAT	2400
QY	2401	GAGGTTCCTCACTCCAGGCCAAACCAAGTACGATTCGACATTTGCCCTCCCAAACTG	2460
Db	2401	GAGGTTCCTCACTCCAGGCCAAACCAAGTACGATTCGACATTTGCCCTCCCAAACTG	2460
QY	2461	GCTCTGCGTGAAGACCCCTGCTCTTTTGCCAGATGACCTTTCTGAGCCCTCCACAACCCC	2520
Db	2461	GCTCTGCGTGAAGACCCCTGCTCTTTTGCCAGATGACCTTTCTGAGCCCTCCACAACCCC	2520
QY	2521	AACCACTGGGCAGACTTCAGTTCCACGTGGCCCAAGCACTCAACGAGAGCCGAAAGC	2580
Db	2521	AACCACTGGGCAGACTTCAGTTCCACGTGGCCCAAGCACTCAACGAGAGCCGAAAGC	2580
QY	2581	GACAACTGGATATAGTGGGCGGCTCAGGCTTCTGACCCGATACCTAGGCTGGGCAAGTTA	2640
Db	2581	GACAACTGGATATAGTGGGCGGCTCAGGCTTCTGACCCGATACCTAGGCTGGGCAAGTTA	2640
QY	2641	CGGCAAGATCAGCTTTTACCACCAAGCCACAGCCACTGGCTCTCCCGATCCCGCTCTG	2700
Db	2641	CGGCAAGATCAGCTTTTACCACCAAGCCACAGCCACTGGCTCTCCCGATCCCGCTCTG	2700
QY	2701	GCGCAGGGTGAAGAAAGTGGGAAGGGCTCAAGCGCAAGCCCTGATTCCTGGAGGCCAAA	2760
Db	2701	GCGCAGGGTGAAGAAAGTGGGAAGGGCTCAAGCGCAAGCCCTGATTCCTGGAGGCCAAA	2760
QY	2761	AAAGACAACAATTAAATTTTAAACAAAAGTACGTCATCAACGTTCTTGGAACAAGAC	2820
Db	2761	AAAGACAACAATTAAATTTTAAACAAAAGTACGTCATCAACGTTCTTGGAACAAGAC	2820
QY	2821	ATGTGTGGTGTGGGAAGTTCAAGGTCAGAGGGTGGTTCCTCCAACTTACGTA	2880
Db	2821	ATGTGTGGTGTGGGAAGTTCAAGGTCAGAGGGTGGTTCCTCCAACTTACGTA	2880
QY	2881	CTCATTTGAGGGCCCGTAAAGGAATTCACAAAGATCGATCTGACCCTTACGAAAGTCT	2940
Db	2881	CTCATTTGAGGGCCCGTAAAGGAATTCACAAAGATCGATCTGACCCTTACGAAAGTCT	2940
QY	2941	GCTAGTCTAAAGAGATGGCTTCCCGGCGCCAGCCAGCCATTTCCCGAAGAGTTT	3000
Db	2941	GCTAGTCTAAAGAGATGGCTTCCCGGCGCCAGCCAGCCATTTCCCGAAGAGAGTTT	3000
QY	3001	ATTGCAATGTACATACGAGAGTCTGAGCAAGAGATTTAACTTTACGCAAGGGAT	3060
Db	3001	ATTGCAATGTACATACGAGAGTCTGAGCAAGAGATTTAACTTTACGCAAGGGAT	3060
QY	3061	GTGATGTGTGTACCAAGAAAGATGTGATCTGTGTGACCGGGAACGATGGGCAAGTCC	3120
Db	3061	GTGATGTGTGTACCAAGAAAGATGTGATCTGTGTGACCGGGAACGATGGGCAAGTCC	3120
QY	3121	GGAGTCTTCCCTTCTAACTATGTAGAGCTTAAAGATTCAGAGGGCTCTGGAACGTGCGG	3180
Db	3121	GGAGTCTTCCCTTCTAACTATGTAGAGCTTAAAGATTCAGAGGGCTCTGGAACGTGCGG	3180
QY	3181	AAAACAGGAGATTAGAAAAAAAACCTGAATTGCCAGGTTATTGCTTCTACGCTCT	3240
Db	3181	AAAACAGGAGATTAGAAAAAAAACCTGAATTGCCAGGTTATTGCTTCTACGCTCT	3240
QY	3241	ACTGGTCCCGAACAACCTACCTCTGGCTCTGGGCAAGCTGATTTCTGATCCGAAAAAGAC	3300
Db	3241	ACTGGTCCCGAACAACCTACCTCTGGCTCTGGGCAAGCTGATTTCTGATCCGAAAAAGAC	3300

QY 841 CTAGCTATGCACTTAATTGATGTGTCATGTCTGTCAGGCACTGCGCCCTGCTGCTT 900
DB 881 CTAGCATGCACTGATGTGATGTGCTGATGTCTGTCAGGCACTGCGCCCTGCTGCTT 940
QY 901 CCGAATAACATCCCTCTCTCTTCAAGAGATTGCTCCGCGCAGTGGAGTGTCCGTAT 960
DB 941 CCGAATAACATCCCTCTCTCTTCAAGAGATTGCTCCGCGCAGTGGAGTGTCCGTAT 1000
QY 961 AGCTCTTCTCTGTCAGTGAAGGCTGCGCAGGAGGCGCTGTCAGAGATGAGCAGCAG 1020
DB 1001 AGCTCTTCTCTGTCAGTGAAGGCTGCGCAGGAGGCGCTGTCAGAGATGAGCAGCAG 1060
QY 1021 CCGAAGAAAGAACTGCTGTCAGTGAAGATTAAGAGCGGAGAACTTCAAGCAGGCG 1080
DB 1061 GTGAAAAGAAAGTGTGCTGTGATTTGAAGATTAAGAGCGGAGAACTTCAAGCAGGCG 1120
QY 1081 AGTGTGAGCTGGAAGAAAGCGCCGCAAGCGCTCTTGAAGCAGCAGCGCAAGCAGAG 1140
DB 1121 AACCTGAGCTGGAAGAAAGCGCCGCAAGCGCTCTTGAAGCAGCAGCGCAAGCAGAG 1180
QY 1141 CGGTTGGCTCAGCTGGAAGCGCGCGCAGCAGAGAGAGAAAGCGGAGCGCCAGAGCAG 1200
DB 1181 CGGCTGCTCAGCTGGAAGCGCGCGCAGCAGAGAGAGAAAGCGGAGCGCCAGAGCAG 1240
QY 1201 GAGGCGAAGCGGAGCTGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
DB 1241 GAGGCGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1300
QY 1261 CAGCGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
DB 1301 CAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1360
QY 1321 GAAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
DB 1361 GAAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1420
QY 1381 AACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
DB 1421 AACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1480
QY 1441 TTGAAGCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1500
DB 1481 TTGAAGCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1540
QY 1501 CCACTGAGCAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560
DB 1541 CCACTGAGCAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1600
QY 1561 GCTGAATACCCCACTTACAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620
DB 1601 GCGGAATACCCCACTTACAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1660
QY 1621 ATTCCAGAGAAACAGATACAGATGATGATGATGATGATGATGATGATGATGAT 1680
DB 1661 ATTCCAGAGAAACAGATACAGATGATGATGATGATGATGATGATGATGATGAT 1720
QY 1681 AAGAGATCGCTTCTTACCTCAAAAGAGCTTGAAGCAAGAGAGAGAGAGAGAGAG 1740
DB 1721 AAGAGATCGCTTCTTACCTCAAAAGAGAGCTTGAAGCAAGAGAGAGAGAGAGAG 1780
QY 1741 CTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1800
DB 1781 CTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1840
QY 1801 GTTTTCAACCAACAGCTGAGAGAACTGAGAGATGATGATGATGATGATGATGATGAT 1860
DB 1841 GTTTTCAACCAACAGCTGAGAGAACTGAGAGATGATGATGATGATGATGATGATGAT 1900
QY 1861 CAGAGATCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1920
DB 1901 CAGAGATCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1960

QY 1921 TTGAAGAGCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1980
DB 1961 TTGAAGAGCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2020
QY 1981 GAGCATGTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2040
DB 2021 GAGCATGTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2080
QY 2041 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2100
DB 2081 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2140
QY 2101 GACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2160
DB 2141 GACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2197
QY 2161 CCCTGCTTACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2220
DB 2198 CCCTGCTTACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2257
QY 2221 GTATATTAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2280
DB 2258 GTATATTAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2317
QY 2281 CAGAGAGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2325
DB 2318 CAGAGAGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2377
QY 2326 GATGAGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2385
DB 2378 GATGAGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2437
QY 2386 AAGATTTCCAGAAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2445
DB 2438 AAGATTTCCAGAAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2497
QY 2446 CCTGCCCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2505
DB 2498 CCTGCCCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2557
QY 2506 CCTGCCCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2565
DB 2558 CCTGCCCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2617
QY 2566 GAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2625
DB 2618 GAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2677
QY 2626 AGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2685
DB 2678 AGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2737
QY 2686 CCATCTCCGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2745
DB 2738 CCATCTCCGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2797
QY 2746 CCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2805
DB 2798 CCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2857
QY 2806 CTGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2865
DB 2858 CTGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2917
QY 2866 AAGCTTTACGTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2925
DB 2918 AAGCTTTACGTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2977
QY 2926 CTTACTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2985
DB 2978 CTTACTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3037
QY 2986 CCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3045

Db	3038	CCTGGAGAGAAATTTGTTGCTATGTACACTTACAGAGATTCTGAGCATGAGATTTTAAC	3097
QY	3046	TTTCAGCAAGGGGATGTGATTTGTGTTACCAAGAAAGTGTGACTGTGACGGGAACG	3105
Db	3098	TTTCAGCAAGGGCATGTGATTTGTGGTTACCAAGAAAGATGTGAGTGTGTGACGGGCACA	3157
QY	3106	GTGGGCCCAAGTCCGAGATCTTCCCTCTTAACCTATGTGAGGCTTAAGATTCAGAGGC	3165
Db	3158	GTGGCCGAGAGTCCGGAGTCTTCCCTCTTAACCTATGTGAGGCTTAAGACTCAGAGGC	3217
QY	3166	TCTGGAATCTGCTGGGAAAAACAGGGAGTTTAAAGAAAAAACCTGAAATTGCCAGGTATT	3225
Db	3218	TCTGGAATCTGCGGGGAAAAACAGGAGTTTAAAGAAAAAACCTGAAATGCCAGGTATT	3277
QY	3226	GCTTCTTACGCTGTACTGTCTCCGAAACACTACCCCTGGCTCTGGGCACTGATTCTG	3285
Db	3278	GCTTCTCTACATGTCTACGGGTCTGTAAACAGCTACCCCTGGCTCTGTCAAGCTGATTTTG	3337
QY	3286	ATCCGGAAAAAGAACCCAGAGTGTGATGTGTGGGAAGAAATCTGCAAGCTGTGAGAAAAAG	3345
Db	3338	ATCCGAAAAAGAACCCCGGTGGAGTGTGGGAAGAAATCTAACAAGCTGTGAGAAAAAG	3397
QY	3346	CGCCAGATAGGGTGGTTTCCAGCAATATGTCAAACTTCTTAAGCCCGGAAACAAGCAA	3405
Db	3398	CGCCAGATAGGGTGGTTTCCAGCAATATGTCAAACTTCTTAAGCCCTGTGAACAAGCAA	3455
QY	3406	ATCAACCCCACTGAGCTACCCAGAAGCCGAGTGCAGCCAGCAGAGTGTGCAGGTGATCGGG	3465
Db	3458	ATCAACCCCACTGAGCTACCCAGAAGCCGAGTGCAGCCAGCAGAGTGTGCAGGTGATCGGG	3515
QY	3466	ATGTACGATTTACACCGCCAGAACGATGACGAACTTACGCTTTCAGCAAAAGCCACATCATC	3525
Db	3518	ATGTACGATTTACACCGCCAGAACGATGACGAACTTACGCTTTCAGCAAAAGCCACATCATC	3577
QY	3526	AACGTCCTCAACAAGAGAGACCCCGGACTGTGGAAAGGAAGTACAGTGGGCAATGTGGG	3585
Db	3578	AACGTCCTCAACAAGAGAGACCCCGGACTGTGGAAAGGAAGTACAGTGGGCAATGTGGG	3633
QY	3586	CTCTTCCCATCCATTATGTAAAGCTGACCAAGACATGTGAACCCAGCAGCAATGA	3642
Db	3638	CTCTTCCCGTCCAAATTATGTGAAGCTGACCAAGACATGTGAACCCAGCAGCAATGA	3694

RESULT 4

AF114488

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

TITLE

JOURNAL

FEATURES

SOURCE

AF114488

Homo sapiens interseectin short isoform (ITSN) mRNA, complete cde.

AF114488

AF114488.1

GI:4808824

5381 bp

mRNA

linear

PRI 16-JUL-2002

5381

Homo sapiens

Homo sapiens (human)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 5381)

Pucharcos, C., Fuentes, J. J., Casaa, C., de la Luna, S., Alcantara, S., Arbones, M. L., Soriano, E., Bectivill, X. and Pritchard, M.

Alu-aplice cloning of human interseectin (ITSN), a putative multivalent binding protein expressed in proliferating and differentiating neurons and overexpressed in Down syndrome

Eur. J. Hum. Genet. 7 (6), 704-712 (1999)

99415290

10482960

2 (bases 1 to 5381)

Pucharcos, C., Fuentes, J. J., Pritchard, M. and Bectivill, X.

Direct Submission

Submitted (16-DEC-1998)

Medical and Molecular Genetics Center, Cancer Research Institute, U/Hospitalat de l'lo., Avia.

Castelldefels km. 2.7, Barcelona 08907, Spain

location/Qualifiers

1..5381

/organism="Homo sapiens"

gene	1. .5381	/gene="ITSN"	269. .3931	/gene="ITSN"	/codon_start=1	/product="Intersectin short isoform"	/protein_id="A02995.1"	/db_xref="GI:4808825"	/translation="MAQPPPTFGSGSLDIWATTVTERANGDOFSLKPEISGETTDQ RNFPGSLDPPVLAQIOMALADNMGDMDOVESIAMKLTKLKOGLOPSPALPPV KQPVATSSAPAFMGGIGASMPILTAAPVAGMPTLVGMSPTLVASPTAAPLAA GAPVITQPLPAPFAHPAATLPKSSFSFSSPESQNTLQKXQSPVAVVPAEVAIV QSSLYKRYQLPNSHDKTMSGHLTGPQATILIMOSLPAQULASINSLDIDODGLTQ LEDFLAWHLIDVAMSGQPLPVLPPETYPPEFRRVSGSGISVISTSDVDRLEPPL LEDOOQLEKLLPTVPEDKKENFERGVLLEKRRQALBOORKEQELQALERAQEK RKREEROBOERKRQLELEKOLEKOROLEEERERKLEEREAKELEEROLEMTE ENRREKROENKROEDIVILKAKKLTLEFELDLERKLEDOIRCLTORR EIBSTNKSRELRILAEITHLOOQLOESQOMLERLIPKQILNDQLKQVQNSIHRSL MEARLBAKELARQLHDLDEVEKETRSKQDEIDIPNNQLKELERINKKOLQOK TLKRLKQEBQERKIIELEKQKEAQRADQRQWELHVOEDBEHQPRKLEHEK KREESVYKKDKGEKQKQADQKGLRFLPHQDEPAPKAPVQAQWMTAEKPLTISAE KVYVRYALVPRESHDEITTOPDQIVWKEKEMDESOTGEPGMGLKGTGTFEP NYAKTIPENEPAPVPKVTBSTSAAPRLALRETPALPVAIVSESTPTNNMADSS WPTSTNEKEPETHNDMAAQAQSLTVPAHQQLRQSAIPTRPATSSPSPVIGQEKV GLQKQALPWRPAKNDNHNFNKNDVITVLEQDDMMRGEEVQGQKMFPSKYVKLISG IRKSTSDMSGSSSPASLKRVAAPAAKPVVSGEEFIATVYESSQGLTQOQGVII VTKKDDQMDWTGTVDKAGVPPSNVYRLDSDSGSAGTGSILGKKPEIAOYASVTA GPEQUTLAAQGLILIRKKNPGCMMWEGLEQARKKQIGMPANVYKULSPGSKITPP EPSPSTLAAACVOVIGWYDYTAQNDDELAFNKQGIINVLNKEDPDMWGEVNGVGL PSNYVXKLTTMDPSQ"
CDS	BASE COUNT	1629 a	1192 c	1304 g	1256 t	ORIGIN			
	Query Match	77.7%;	Score 2831;	DB 9;	Length 5381;				
	Best Local Similarity	86.5%;	Pred. No. 0;						
	Matches 3167;	Conservative	0;	Mismatches 475;	Indels 21;	Gaps 3			
Oy	1	ATGGCTCAGTTTCCACACCTTTGGGTGATGACCTGATGTCGTGGGCATTAATCTGTGAG	60						
Db	269	ATGGCTCAGTTTCCACACCTTTGGGTGATGACCTGATGTCGTGGGCATTAATCTGTGAG	328						
Oy	61	GAAAGGCCAAGCATGACACAGCAGTCTTGTGCGTGAAGCCGATAGCGGAGATTATTCT	120						
Db	329	GAAAGAGGGAAGCATGATCAGACAGTCCATAGTTTAAAGCCAAATATCTGGATTCAATACT	388						
Oy	121	GGTATCAAGAGGAGAACTTTTTCATCTGGGTGTAACCTCAAGCCTGTCTTAGCACAA	180						
Db	389	GGTATCAAGCTAAGAACTTTTTCATCTGGGTGTAACCTCAAGCCTGTCTTAGCACAG	448						
Oy	181	ATATGGGCGCTAGCGGACATGATATACGATGGAAGATGATCAAGTGAATTTTCCATA	240						
Db	449	ATATGGGCACTAGCTGACATGAATATGATGGAAGAAATGATCAAGTGAAGTTTCCATA	508						
Oy	241	GCCATGAAAGCTTATCAAACTGAAGTATAAAGATATACAGCTCCCTCCACACTTCCCT	300						
Db	509	GCTATGAAACTTATCAAACTGAAGTATAAAGATATACAGCTCCCTCGACACTTCCCT	568						
Oy	301	GTCATGAAACAGCAACAGTGGCTATTTCCAGTGCACAGCATTTGTATAGAGAGGATT	360						
Db	569	GTCATGAAACAGCAACAGTGGCTATTTCTAGCGCACAGCATTTGGTATGGAGGATATC	628						
Oy	361	GCTAGCATGCAACCACTACAGCTGTGCTCTGTGCGCAATGGGCTCCATTTCAAGTTGT	420						
Db	629	GCCAGCATGCCACCGCTTACAGCTGTGCTCCAGTGCACATGGGATTCATTCAGTTGT	688						
Oy	421	GGATATGCTCAACCTTATGATATCTGTGCTCCTCCAGAGACAGATGCTCCCTGGCTAAC	480						
Db	689	GGATATGCTCAACCTTATGATATCTGTGCTCCTCCAGACAGAGCTGTGCTCCCTGGCTAAC	748						

OY	481	333GGGCTCTCCCGTACACAGCTTGTGCTGCTTGGCATCTCTGACGCCACATGGCCA	540
Db	749	GGGGCTCCCCCTGTATACACCTTCGCTGTGATTTGTCACTCTGCAGCCACATTGGCA	808
OY	541	AAGAGTTCTTCCTTACAGAGATCTGGTCCAGGGTCACAATTAAACCTAACGTTACAGAG	600
Db	809	AAGAGTTCTTCCTTATAGTAGATCTGGTCCAGGGTCCACATTAACACTTAAATTACAAAG	868
OY	601	GCACAACTCATTCGATGTGCGCAGCGCCCTCCAGCAGCAGAGATGGGCTGTGCTCAGTCA	660
Db	869	GCACAGCTATTGATGTGGCCAGTGTCCCAACAGTGGCAGATGGGCTGTTCTCTCAGTCA	928
OY	661	TCAAGGCTGAAATACAGCAGTATTCAACAGCCACGACAAACTATATAGTGGACACTTA	720
Db	929	TCMAACCTGAAATACAGCCAATTATTCAAATGTCTATGACMAAACTATAGTGGACACTTA	988
OY	721	ACAGGTCCCCAGGCGCAAGAACTATTCTCANGCAATCAATTAACTCCAGGCTCAGTGGCT	780
Db	989	ACAGGTCCCCAGCAGAACTATTCTTATGCGTAACTTAACTTAACTCCAGGCTCAGTGGCT	1048
OY	781	TCAATATGGAATCTTTTCTGACATTTGATCAAGATGGAAAACTCACTGCAGAAAGAAATTATC	840
Db	1049	TCAATATGGAATCTTTTCTGACATTTGATCAAGATGGAAAACTTACAGCAGAGAGAAATTATC	1108
OY	841	CTAGCTATGCACTATTGATGTTGGCCATGTTCTGGTCAAGCCATCGCCCGCTCTGCT	900
Db	1109	CTGGCAATGCACTCATTTGATGTTAGCTTGTCTGGCCAAACCACTGCTCCTGCT	1168
OY	901	CCAGAAATCATCCCTCTCTCTTCAGAGATTTGGCTCCCGCAGTGGGATCTCCGTACTA	960
Db	1169	CCAGAAATCATCTCACTCTTTTGAAGAGTTGCATCTGGCAGTGGTATATCTGTCA	1228
OY	961	AGCTCTTCTTGTGATCAGAGGCTGCTGAGGACCGTGTCAAGAGATGAGCAGAG	1020
Db	1229	AGCTCAATCTGTAGATCAGAGGCTTACAGAGAACCAAGTTTATGAAGATGAACACAA	1288
OY	1021	C---CAGAGAAAGAACTGCTCTGTGACATTTAGATTAAGAACCGGAGAACTTGCAGCA	1077
Db	1289	CAATTAGAAAAGAAATTACTGTGAACGTTTGAAGATAAGAACCGGAGAACTTGAACGT	1348
OY	1078	GGCAGTGTGAGCTGTGAGAGAGCGCCGCAACGCTCTTGGAGCAGACGGCAAGAGAG	1137
Db	1349	GGCAACTGTGAATCTGAGAAACGAAAGGCAACTCTCTCTGAAACAGCAACGCAAGAGAG	1408
OY	1138	GAGCGGTTGGCTCAGCTGAGCGCGCCGAGAGAGAGAAAGAACCGGAGCGCCAGGAG	1197
Db	1409	GAGCGGCTTGGCTCAGCTGAGCGCGCGGAGAGAGAGAGAGAGAGCGCTGAGCGCCAGGAG	1468
OY	1198	CAGAGGCTCAGCGCAGCTGTGAGAGCAGCTGTGAAGAACAGCGGAGCTGTGAG	1257
Db	1469	CAAGAGCGCAAAAGCACTGTGAATCTGTGAAGCACTGTGAAGAACAGCGGAGCTGTGA	1528
OY	1258	CGGAGCGCAGAGAGAGAGAGAGAGAGATCCAGAGCGCCGAGCGCCGCAAAACGGGAA	1317
Db	1529	CGGAGAGAGAGAGAGAGAGAGAGAAATTGAGAGCGCAGAGCGCTGCAAAAACGGGAA	1588
OY	1318	CTGGAAGGCAAGCAGCACTTGAATGGGAAACGGAACCGGAGACAGAACTCTCTGAATCAG	1377
Db	1589	CTTGAAGGCAAGCAGCACTTGAATGGGAAACGGAATCGAAGGCAAGAACTCTTAATCA	1648
OY	1378	AGGAACAGAGAGCAGAGAGGGAACCGTGTCTGTAAAGCAGAGAGAACTCTGTGAGTTT	1437
Db	1649	AGGAACAGAGAGCAAGAGGAGCATAGTTGTACTGAAGCAAGAAAGAAAGACTTGTGAAATTT	1708
OY	1438	GAGTTAGAACTTGTGAATGACAAAAGATCTACGTAAGAAAGAAACTTACAGATATCAG	1497
Db	1709	GAGTTAGAACTTGAATGATTAAGATTAAGATCACTTGAAGAGGAACTTCAAGATATCAG	1768
OY	1498	TGTGACATGCGCAACCCAGAGCCAGAAATTTAGAGACGAAACAAGTCTAGAGACTTAGA	1557
Db	1769	TGTGACATTTGACCAACCCAGAGCCAGAAATTTAGAGACCAAAACAATCTTAGAGATTAGA	1828
OY	1558	ATTGCTGAATTCACCACTTACAGCAGAGTGTGACAGAACTCTCAGCAAACTCTTGGAGA	1617

Db	1829	ATTGCCGAAATCACCCATCTTACACAACTAATTAACGGAATTCAGCAATCTTGAGAGA	1888
OY	1618	CTTATTCAGAGAAACAGATACATCAGAGACCAAGTTAAAAAAGATCCAGCAACAGTTTG	1677
Db	1889	CTTATTCAGAAAAACAGATCTCAATGACCAATTAAAAAAGATTACAGACAAACAGTTTG	1948
OY	1678	CATAGAGACTCGCTTCTTACCCCTCAAAAGACCTTGGAAAGCAAGAGAGCTGGCCCGCAG	1737
Db	1949	CACAGAGATTCACTTGTGTTACCTTAAAAAGACCTTTAGAGCAAAAAAGAACTAGCTGGCAG	2008
OY	1738	CAGCTCCGGAGACAGCTGACGAGGTGAGAGAGAACCAAGTCAAACTGCAGAGATT	1797
Db	2009	CACCTACGAGACCAACTGAGTGAAGTGGAGAAAGAACTAGATCAAACTACAGAGAGATT	2068
OY	1798	GATGTTTCAACAACCAAGCTGAAGAACTGAGAGATACATTAACCAACACAACTCCAG	1857
Db	2069	GATATTTTCAATTAATCAGCTGABAGAACTAAGAAATACATTAAGCAACCACTCCAG	2128
OY	1858	AAGCAGAGSTCCCTGAGAGCCGACCTGAAAGCAGAAAGAGACGAGAGAGAGAGCTG	1917
Db	2129	AAGCAAAAGTCCATGAGAGGCTGAAGAGACTGAACAGAAAGAACAGAAAGATCATA	2188
OY	1918	GAGTTAGAGAACCAAAAGAAAGACGCTCAGAGACGAGTTCAAGAAAGGACCAACATGG	1977
Db	2189	GAAATTAGAAAAACAAAAAGAAAGAGCCCAAGACGACTCAGAAAGGAGCAACAGATGG	2248
OY	1978	CTGAGAGCATGTCAGCAGGAG--GACGAGCAAGCCCCCGAATCCCAACGAGAGAGAC	2034
Db	2249	CTGAGCATGTGTCAGCAGAGAGACGAGATTCAGAGCCAAAGAAATCTTCAAGAAAGAGA	2308
OY	2035	AGACTGAAGAGGAGAGACAGTGTCAAGAGAAAGAGCGGAGAGAGAGCCGCAACCCGAA	2094
Db	2309	AAACTGAAAAAGGAGAGAGAGTGTCAAAAAAGAGATGGCAGAGAAAAAGCAACAGAA	2368
OY	2095	ATGCAAGACAGCAGAGTGGCTTTTCCATCCGATCAAGAGCCAGCTAAGCTGGCAAC	2154
Db	2369	GCACAAGACAGCTGGGTGGCTTTTCCATCAACACCAAGAACAGCTTAAGCCAGTGTCT	2428
OY	2155	CAGGCAACCTGGTCTTCAACAAGAAAGAGCCCGCTTACATTTTGCACAGAGAAAGTGA	2214
Db	2429	CAGGCAACCTGGTCTTCAACAAGAAAGAGTCACTTACATTTTGCACAGAGAAAGTGA	2488
OY	2215	AAAGTGTATTTATCCAGACGCTGTACCCCTTTGAATTCAGAAATCAAGATGATCAAC	2274
Db	2489	AAAGTGTGTATTTACGGGCACTGTACCCCTTTGAATTCAGAAATCAAGATGATCACT	2548
OY	2275	ATCCAGCCAGAGATATATAGTCAT-----GGTGAATGAAGCCAGACTGGA	2319
Db	2549	ATCCAGCCAGAGACATATAGTCATGTTTAAAGGGGAATGGGTGATGAAGCCAACTGGA	2608
OY	2330	GAGCCAGAGATGGCTTGGAGAGAGACTGAAGAGGAAGACGGATGTTTCCCTGCACATAT	2379
Db	2609	GAACTCCGGCTGGCTTGGAGAGAAATTAAGAAAGAAACAGGGTGGTTTCCCTGCACATAT	2668
OY	2380	GCAGAAAAAGATTCAGAAAAATGAGGTTCCCACTCCAGCAAAACAGTGAACGATCTGACA	2439
Db	2669	GCAGAGAAAAATCCAGAAAAATGAGGTTCCCGCTCCAGTGAACCAAGTACATGATTCACA	2728
OY	2440	TCTGCCCCCTGCCCCAACTGGCTCTGCGTAGACCCCTGCTCTTTGCCAGTGAACCTT	2499
Db	2729	TCTGCCCCCTGCCCCAACTGGCTCTGCGTAGACCCCGCCCTTTGGCAGTAACTCT	2788
OY	2500	TCTGAGCCCTTCCAAACCCCAACAATGGGCAAGCTTCAAGTTCCAGCTGGCCCAAGAC	2559
Db	2789	TCAGAGCCCTTCCAAAGCCCTTAAATACCTGGCCCACTTCAAGCTTCCAGCTGGCCCAAGAC	2848
OY	2560	TCAACAGAGAAAGCCAGAAACGAGCAACTGGGATACGTGGAGCGCTCAGACCTTCTTGACC	2619
Db	2849	ACGATATGAGAAACCAAGAAACGAGATACCTGGGATGCAATGGCAGCCCAAGCCCTCTCAGC	2908
OY	2620	GTACCTAGTGTGGCCAGTTTACGGCAGAGATCAGCTTTTACCCAGCCACAGCCACTGGC	2679

Db	2909	GTTCGAAGTGC	CGGCACTTAAGCAGAGAGTCCGCTTTACTCGAGCAGCCGCACTGGC	2966
Qy	2680	TCCTCCCATCT	CCCGCTCTGGGCGAGGGTGAAGGTGGAAAGGGCTACAGCGCAAGC	2735
Db	2969	TCCTCCCGGTCT	CCCTGTCTAGGCCAGGGTGAAGGTGGAGGGGCTCAAGCTCAAGCC	3028
Qy	2740	CTGTATCCCTG	AGAGCCAAAAAGACAACAACCTTAATTTTAAAGTAAGTCACTCATC	2799
Db	3029	CTATATCTTGG	AGAGCCAAAAAGACAACAACCTTAATTTTAAAGTAAGTCACTCATC	3088
Qy	2800	ACCGTCTCGAA	CAGCAAGACATGTGTGTGGAGAGTTCAAGTCAAGAGGGTTGG	2855
Db	3089	ACCGTCTCGAA	CAGCAAGACATGTGTGTGGAGAGTTCAAGTCAAGAGGGTTGG	3148
Qy	2680	TTCCCGAAGT	CTTAAGTAATCTAATTCAGGGCCCGTAAGGAATTCACAAAGATTCAT	2915
Db	3149	TTCCCGAAGT	CTTAAGTAATCTAATTCAGGGCCCGTAAGGAATTCACAAAGATTCAT	3208
Qy	2920	ACTGGCCCTAC	TGAAAGTCTGTGACTTAAGAGAGTGGCTCCCGGCGCCAGACCA	2977
Db	3209	TCGTGTTCTT	CAGAGAGTCTGTGACTTAAGAGAGTGGCTCCCGGCGCCAGACCG	3266
Qy	2980	GCCATTCCCG	AGAAAGATTATGGCATGTACATACAGAGTTCTGAGCAGAGAT	3033
Db	3269	GTCTGTTCCG	GAGAAATTAATGCCATGTACACTTCAGAGTTCTGAGCAGAGAT	3328
Qy	3040	TTAACCTTTCA	GCAAGGGGATGTATGTGGTTACCAAGAAAGATGTGACTGTGGAGAG	3099
Db	3329	TTAACCTTTCA	GCAAGGGGATGTATGTGGTTACCAAGAAAGATGTGACTGTGGAGAG	3388
Qy	3100	GGAACGATGG	CGACAAAGTCCGGAATCTTCCCTTAACTATGTAGAGCTTAAGATTCA	3155
Db	3389	GGAACGATGG	CGACAAAGTCCGGAATCTTCCCTTAACTATGTAGAGCTTAAGATTCA	3448
Qy	3160	GAGGGCTCTG	GAACCTGTGGGAAACAAGGAGTTTAGAAAAAAACCTGAATTGCGCAG	3212
Db	3449	GAGGGCTCTG	GAACCTGTGGGAAACAAGGAGTTTAGAAAAAAACCTGAATTGCGCAG	3508
Qy	3220	GTTATTTGCTT	CCTACGCTGCTACTGGTCCCGAACAATCAACCTGAGCTCTTGGGAGCTG	3273
Db	3509	GTTATTTGCTT	CCTACGCTGCTACTGGTCCCGAACAATCAACCTGAGCTCTTGGGAGCTG	3566
Qy	3280	ATTCTGATCC	GGAAGAAACCAGGTGATGTGGAGAGAGAACTGACAGCTGAGAGG	3333
Db	3569	ATTCTGATCC	GGAAGAAACCAGGTGATGTGGAGAGAGAACTGACAGCTGAGAGG	3628
Qy	3340	AAAAAGGCGC	AGATGGGTTCACAGAAATTATGTCAAACTTCAAGCCCGGGAACA	3399
Db	3629	AAAAAGGCGC	AGATGGGTTCACAGAAATTATGTCAAACTTCAAGCCCGTGGAGCG	3688
Qy	3400	AGCAAAATCAC	CCCCCACTGAGCTAACCAAGACCGGAGTGCAGCCAGAGTGTGCCAGGTG	3455
Db	3689	AGCAAAATCAC	CCCCCACTGAGCTAACCAAGACCGGAGTGCAGCCAGAGTGTGCCAGGTG	3748
Qy	3460	ATCGGAGATGA	CTTAACACCGCCAGAACATGACGAATCTAGCTTACAGCAAGCCAG	3513
Db	3749	ATCGGAGATGA	CTTAACACCGCCAGAACATGACGAATCTAGCTTACAGCAAGCCAG	3808
Qy	3520	ATCATCAACGT	CTCAACAGGAGACCGGACTGTGAGAAAGAGAAAGTCAGTGGGCA	3573
Db	3809	ATCATCAACGT	CTCAACAGGAGACCGGACTGTGAGAAAGAGAAAGTCAATGAGCA	3866
Qy	3580	GTTGGGCTCTT	CCATCCAAATTATGTAAAGCTGACCAAGACATGAGCCCGCAGCA	3633
Db	3869	GTTGGGCTCTT	CCATCCCAATTATGTAAAGCTGACCAAGACATGAGCCCGCAGCA	3928
Qy	3640	TGA	3642	
Db	3929	TGA	3931	

LOCUS	AF114487	6439 bp	RNA	linear	PRI 16-JUL-2002
DEFINITION	Homo sapiens interectin long isoform (ITSN) mRNA, complete cds.				
ACCESSION	AF114487				
VERSION	AF114487.1	GI:4808822			
KEYWORDS					
ORGANISM	Homo sapiens (human)				
SOURCE	Homo sapiens				
REFERENCE	Eukaryotes; Metazoa; Chordata; Craniata; Vertebrata; Euleleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
AUTHORS	1 (bases 1 to 6439) Pucharcos,C., Fuentes,J.J., Caas,C., de la Luna,S., Alcantara,S., Adobes,M.T., Soriano,E., Estivill,X. and Pritchard,M.				
TITLE	Alu-splice cloning of human interectin (ITSN), a putative multivalent binding protein expressed in proliferating and differentiating neurons and overexpressed in Down syndrome				
JOURNAL	Eur. J. Hum. Genet. 7 (6), 704-712 (1999)				
MEDLINE	99415290				
PUBMED	10482960				
REFERENCE	2 (bases 1 to 6439) Pucharcos,C., Fuentes,J.J., Pritchard,M. and Estivill,X.				
AUTHORS	Direct Submission				
TITLE	Submitted (16-DEC-1998) Medical and Molecular Genetics Center, Cancer Research Institute, L'Hospitalet de Llob., Avia.				
JOURNAL	Castelldefels Km. 2,7, Barcelona 08907, Spain				
FEATURES	Location/Qualifiers				
SOURCE	1..6439				
gene	/organism="Homo sapiens"				
CDS	/mol_type="mRNA"				
	/db_xref="taxon:9606"				
	/chromosome="21"				
	/map="21q22.1-q22.2"				
	1..6439				
	/gene="ITSN"				
	269..5434				
	/gene="ITSN"				
	/codon_start=1				
	/product="Intersectin long isoform"				
	/protein_id="A029952.1"				
	/db_xref="GI:4808823"				
	/translation="MAQFPPGSGSLDIMALITVEERKNDQOFSLKPIGSLFGDQAA RNPFGSGLPQPIALQIMALDNGRMQVRSIAMKILKLGQYGLPALPVMW KOQPAISSAPARFGMGSIAMPPLTAVPMSIPGMSPTLQSVPAAPPLANN GAPVIOPLPAPHPAPLALTPKSSFSFGSGSQNLTKQAQSFDAVPPVEMAVP QSSRLAKROLFNSHDKTMSGLTGPQARTILMQSSLQQAOLASITWSDIDQGLTA EERFLAMHLIDVMSGQPLPVPVPEVITPSPFRVRSGSIVSISTVDQRLPEEVP LEBEOQLKELKLVPTFEDKRENFERNLELEKRRQLLQQRKEQRLALEAEOB RKREBOERKRQLELEKOLEKRELERBERREKREIRERAKELEROLREBER RNRQELINQNRKEQDPIVYLKAKKLTPELALNKKHOLESKLODICEQLTQGO EISTNSRELRLAIEITHLQAOLOESQQLGRILPEQQLINDQKQYQNSLRDSIV TLRRAEKELRLAHLHDLDQDEVEKSTRSLQETIDFNOLKEIRETHNQLOLQKQS MEBRALQKEOERKIIELEKQKEQAQERQABRKQWLEHYQOEDHQRPRLHEEKL KRRESYKDKQEGKEQKQADQKGLRFLHQEGKPAKPVQAQPMWTAEGPLTIAQENVN KNVYRALVPFESSHDEITIQPDIVMWKGEWDESGEPGLGGLKGTMPFAPA NVAEKLPENEVPAIPVKYVMTDSTAPAPKILARETPALVATSSPPTPNMADFSST WPSTNKEPBTDMWMAAPSLUTVPSAGQLRQSAATPATATCSSPSPULOGEKRE GLQAOALVPRAKKDNHLNFKNDVITVLQOQNMWGEVQOQGMFPKSYVLIKISRP IRKSTSDSSGSSPALKRNPVPAAPKPVVSGEEFLAMTYESSQGLTFFQOQDVLV VTKDDGMWMTGTQDKAGVPPSNVYRLKDSGEGTAGKTSGLKKRPKIAQVITAN GPBOLLAPQQLILRKKNPGMWEGELQARKKRQIGMFPANVYKLISGTSKITPTPT EPPSTLAIVCAQVITGMVDTAQNDDFLANRKOQIIVLNKKEDPDMMGEWGVGLPE PSNVYKLTDMDSQOQCDLHLDMTLTPERRKQGYIHILVTEENVYNDLQVTEIL FOYKLESELITTEKEVAMITVNMKELIWCILIKLALPRKKSSEGMPPMTDIIIS AOLPMPQPIYRFSRQLNAGALLQOKTDEADPEFEKRIALMPDRCKGMLPSTFLRP MORTAPPLIKNILENTPENHPDHSLKALAEALQVNEGVAKENSRLMEI QAHVQCGLSQEQVFNVTNGLCPKFLKPELIVYKAKNSDELGFVFNEDLLLOITRK PLGSGDKVSPKSNLOYMWKTKPIPLNVLVGLKPNSGDGEIPIHSHIDRYLTR AESINETNAPVOKITKASELYITEKKRERKAYLVSRQOQVGRMLVNMVEGIELRP CRHGKSNPYCEVTMGSQCHITTIQDTPLNKNSNQFPIRDEQVCLCITYERQD PSPDPLGRLEINADIKQQSGVPTKCLLHAETVGEIVARLDQLQDPEP				

QY	2155	AAGCACCCTGAGTCTACCA	CAGAGAAAGACCGCGTTC	ATCATTTCTGCACAGAGAGTGA	221.4			
Db	2429	CAGGCACTTGTATCTG	CACTGCAGAAAAGGTCA	CTTACATTTTCTGCACAGAAAATGA	2488			
QY	2215	AAAGTGTATATTAT	CCGAGCGCTGTACCC	CTTTGAATCAGAAAGTCA	CGATGAGATCAC	227.4		
Db	2489	AAAGTGTATATTAC	CGGGCACTGTACCC	CTTTGAATCAGAAAGTCA	GATGAAATCACT	254.8		
QY	2275	ATCCAGCCAGGAGATAT	AGTCTAT-----	GGTGAATGAAGGCA	CTGGA	231.9		
Db	2549	ATCCAGCCAGGAGATAT	AGTCTAT	AAAGGGAGATGGGTGA	ATGAAGCCAAC	CTGGA	260.8	
QY	2320	GAGCCAGGATGGCTTG	AGAGAGAGTGAAGG	GAAGCGGATGGTTC	CTGCATACTAT	237.9		
Db	2609	GACCCCGCTGGCTTG	AGAGAGAGATTA	AAAGGAAGACAGG	TGGTTCCTGCATACTAT	266.8		
QY	2380	GCAGAAAAGATTC	CAAAAAATGAGTTC	CACTCCAGCCAAAC	CAATGACGATCTGCA	243.9		
Db	2669	GCAGGAAAATTC	CAAAAAATGAGTTC	CACTCCAGGAAAC	CAATGATCTGATCA	272.8		
QY	2440	TCGTGCCCCGCCCC	CAAACTGGCTCTG	GTGAGACCCCGTCC	CTTTCAGTGA	CTCT	249.9	
Db	2729	TCGTGCCCCGCCCC	CAAACTGGCTCTG	GTGAGACCCCGTCC	CTTTCAGTGA	CTCT	278.8	
QY	2500	TCGTAGCCCTTC	CAACCCCAAC	CACTGGCA	CACTTCA	CGTGGCCAGCAGC	255.9	
Db	2789	TCAGAGCCCTTC	CAACCCCAAC	CACTGGCA	CACTTCA	CGTGGCCAGCAGC	284.8	
QY	2560	TCAAACAGGAAG	CCAGAAACCGCA	CACTGGATATAG	TGGGGGGCTCA	CGCTCTCTGAC	261.9	
Db	2849	ACGAATGAGAA	CCAGAAACGATATAC	TGGGATGATGGG	CAAGCCCA	CGCTCTCTCA	290.8	
QY	2620	GTACTAGTGTGCTG	GCAGATTAACG	CAAGATCAGCCTTAC	CCCAAGCA	CAAGCACTGGC	267.9	
Db	2909	GTTCAGATGTC	CGCGCACTTAAG	CAAGATGTC	CGCTTATCTCA	CAAGCAAGGCACTGGC	296.8	
QY	2680	TCCTGCCCATCT	CCCCCTCTGGG	CCAGGCTGA	AAAGGTGA	AGGCTTCA	AGCCAGCC	273.9
Db	2969	TCCTGCCCATCT	CCCCCTCTGGG	CCAGGCTGA	AAAGGTGA	AGGCTTCA	AGCCAGCC	302.8
QY	2740	CTGTATCCCTG	AGAGCCAAAA	AGCAACCACTTA	ATTTTAA	CAAAATGATGATCTATC	279.9	
Db	3029	CTATATCTCTT	AGAGCCAAAA	AGCAACCACTTA	ATTTTAA	CAAAATGATGATCTATC	308.8	
QY	2800	ACCGTTCGGA	CAAGCAAGATGTG	GTGTGGAGAGTT	CAAGGTCA	AGAGGGTTGG	285.9	
Db	3089	ACCGTTCGGA	CAAGCAAGATGTG	GTGTGGAGAGTT	CAAGGTCA	AGAGGGTTGG	314.8	
QY	2860	TTTCCCAAGTCTTA	AGTAACTCATTT	CAAGGCCCCGTA	AAAGAAATCC	CAAGCATCGAT	291.9	
Db	3149	TTTCCCAAGTCTTA	AGTAACTCATTT	CAAGGCCCCGTA	AAAGAAATCC	CAAGCATCGAT	320.8	
QY	2920	ACTGGCCCTACT	GAAAGTCTGCTAGT	CTAAAGAGAGTG	CGCTCCCGCGCC	CAAGCA	297.9	
Db	3209	TCGTGTTCTT	CAGAGAGTCTGCTAGT	CTTAAGGAGATAG	CTCTCCAGCA	AGCCAGCG	326.8	
QY	2980	GCCATTTCCCG	GAAGAGTTATTG	CCATGTACATAC	AGAGTTCTTGA	CGAGAGAT	303.9	
Db	3269	GTCGTTTCGGG	GAAGATTTATTG	CCATGTACATAC	AGAGTTCTTGA	CGAGAGAT	332.8	
QY	3040	TTAACCTTTAC	AGAGGAGATGTAT	GTGGTTTAC	CAAGAAAGATGTGA	CTGGTGAGG	309.9	
Db	3329	TTAACCTTTAC	AGAGGAGATGTAT	GTGGTTTAC	CAAGAAAGATGTGA	CTGGTGAGG	338.8	
QY	3100	GGAACGATGGG	CGACATGTCGAGTCT	TCCTTTCTAA	CTATGTGAGGCTTAA	AGATTTCA	315.9	
Db	3389	GGAACGATGGG	CGACATGTCGAGTCT	TCCTTTCTAA	CTATGTGAGGCTTAA	AGATTTCA	344.8	
QY	3160	GAGGCTCTTGA	ACTGCTGGGAAA	CAGGAGTTT	AGAAAAAA	CTGGAATTTGCCAG	321.9	
Db	3449	GAGGCTCTTGA	ACTGCTGGGAAA	CAGGAGTTT	AGAAAAAA	CTGGAATTTGCCAG	350.8	

QY	3220	GTAAATTGTTCTTACGCTGCTACTGATCGCCGAACAATCAACCTCGGCTCCCTGGCAGACTG	3279
Db	3509	GTAAATTGCTTATACACCGCCACCGGCCCCGAGACGCTCACTCTCGCCCTGTGTCACTG	3568
QY	3280	ATTCTGATTCGGAAAAAGAACCCAGTGATGTGGGAAGAGAACTGCCAAGCTTCGAGG	3339
Db	3569	ATTTTGATTCGGAAAAAGAACCCAGTGATGTGGGAAGAGAGCTGCAAGCACTGTGGG	3628
QY	3340	AAAAAGGCCAGATAGGATGGTTCCAGCAAAATTAATGCAAACTTAAAGCCCGGAA	3399
Db	3629	AAAAAGGCCAGATAGGATGGTTCCAGCTTAATATGTAAGCTTCTTAAGCCTTGGAGC	3688
QY	3400	AGCAAAATCAACCCCACTGAGCTACCCAGAACCGCAGATGACGCCAGCAGTGTCCAGGTG	3459
Db	3689	AGCAAAATCACTCAACAGAGCCACTTAATGCAACAGCAATTAGCGGCAGTGTCCAGGTG	3748
QY	3460	ATCGGAGATGTAACGATTTACACCGCCCAAGACGATGACAACTAGCTTCAAGCAAAAGGCCAG	3519
Db	3749	ATTGGGATGTACGACTACACCGCCCAAAATGACATAGCTGGGCTTCAACAAAGGCCAG	3808
QY	3520	ATCATCAACGCTCTCAACAAGAGAGACCCGGACTGTGTGAAGAGAGAAAGTCACTGGGCCAA	3579
Db	3809	ATCATCAACGCTCTCAACAAGAGAGACCTGTGACTGTGTGAAGAGAGAAAGTCAATGACAA	3868
QY	3580	GTTGGGCTCTTCCCATCCAAATTAATGTAAGGTGACCAAGACATGACCCCAAGCCAGCA	3639
Db	3869	GTTGGGCTCTTCCCATCCAAATTAATGTAAGGTGACCAAGACATGACCCCAAGCCAGCA	3928
QY	3640	TG 3641	
Db	3929	TG 3930	

```

RESULT 6
AF064243
LOCUS AF064243 5287 bp mRNA linear PRI 21-NOV-1998
DEFINITION Homo sapiens intersectin short form mRNA, complete cds.
ACCESSION AF064243
VERSION AF064243.1 GI:3859852
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 5287)
2 (bases 1 to 5287)
Guipponi,M., Scott,H.S., Chen,H., Schebesta,A., Rossier,C. and
Antonarakis,S.E.
Two isoforms of a human Intersectin (ITSN) protein are produced by
brain-specific alternative splicing in a stop codon
Genomics 53 (3), 369-376 (1998)
99017974
MEDLINE
9799604
PUBMED
2 (bases 1 to 5287)
Guipponi,M., Scott,H.S., Chen,H., Schebesta,A., Rossier,C. and
Antonarakis,S.E.
Direct Submission
Submitted (05-MAY-1998) Genetics and Microbiology, CMU, 1 rue
Michel-Servet, Geneva 4 CH-1211, Switzerland
FEATURES
SOURCE
1..5287
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.1-q22.2"
/tissue_type="brain"
/dev_stage="fetus"
107..3769
/codon_start=1
/product="intersectin short form"
/protein_id="AAC78610.1"
/db_xref="GI:3859853"
/translation="MAQFPFPGGSLDIWAIYVERAKHQDQFSLKPIGSPITGDQA

```


[illegible]

```
misc_feature      167. .406 /note="encodes EH domain"
misc_feature      767. .1936 /note="encodes EH domain"
misc_feature      2324. .2524 /note="encodes SH3 domain"
misc_feature      2843. .3019 /note="encodes SH3 domain"
misc_feature      3110. .3286 /note="encodes SH3 domain"
misc_feature      3326. .3520 /note="encodes SH3 domain"
misc_feature      3569. .3748 /note="encodes SH3 domain"
misc_feature      /note="encodes SH3 domain"

BASE COUNT      1635 a 1169 c 1227 g 1256 t
ORIGIN
```

QY	481	GGGGGCTCTCCCGGTCATACAGCCCTGCGCTGCGGTTGGGCACTCTGACGCCAATGGCA	540
Db	587	GGGGCTCCCCCTGTATCAACCTCTGCTGCATTTGGTCACTCTGGACCAATGGCA	646
QY	541	AAGATTCTTCTTCAGCAGATCTGTCTCAGGGTCACAATTAAACCTAAGTTACAGAA	600
Db	647	AAGATTCTTCTTAGTAGATCTGTCTCAGGGTCACACTAAACCTAATTACAAAG	706
QY	601	GCACATATTCGATGTCTGCCAGGGCCCTCAGCAGCAGATGGGCTGTGCTCACTCA	660
Db	707	GCACAGTATTGTATGTGGCCAGGTCTCCACAGATGGCAGGTGGGCTGTCTCACTCA	766
QY	661	TCAAGGCTGAATACAGGCACTATTCAAGCCACGCAAAACATATAGTGGACATCTA	720
Db	767	TCAAGCTGAATACAGGCACTATTCAATAGTCAATACAAACTATAGTGGACATCTA	826
QY	721	ACAGGTCCCAAGGCAAGACTATTCATGCAATCAAGTTATCCCAAGGCTCAAGCTGCT	780
Db	827	ACAGGTCCCAAGGCAAGACTATTCATGCAATCAAGTTATCCCAAGGCTCAAGCTGCT	886
QY	781	TCAATATGGAATCTTTTCGATGATTCATAATGGAATCTCACTGCAAGAAATTTATC	840
Db	887	TCAATATGGAATCTTTTCGATGATTCATAATGGAATCTTTAAGCAAGAAATTTATC	946
QY	841	CTAGCTATGACCTAATATTGATGTGCCATGTCTGTGACGACATCGCGCCGTCTGCT	900
Db	947	CTGGCAATGACCTCAATTGATGTGATATGTCTGGCCAAACCATCTGACCATCTGTCTGCT	1006
QY	901	CCAGAAATCATCTCTCTCTCTTCTTCAAGAAAGTTGCTCGGCAAGTGGATTCCTGCTATA	960
Db	1007	CCAGAAATCATCTCTCTCTTCTTCAAGAAAGTTGATCTGGCAAGTGGATTCCTGCTATA	1066
QY	961	AGCTCTTTCTGTGTGATCAGAGGCTGCTCAGAGCGCGTGTGAGAGATGAGACAGAG	1020
Db	1067	AGCTCAACATCTGTAGATCAGAGGCTCACAGAGGAACAGTTTAGAAGATGAACAA	1126
QY	1021	C---CAGGAAGAACTGCTGTGACATTTGAAATGAAGACGGGGAACCTTGCAGCA	1077
Db	1127	CAATTAGAAAGAAATTAACGTGAACGTTTGAATAAAGACGGAGAACTTTGAACGT	1186
QY	1078	GGCAGTGTGAGCTGAGAGAACGCCCGCCAGACGCTCTTGGACGACAGCCGAAGACAG	1137
Db	1187	GGCAACCTGAGAACTGAGAGAAACGAAGGCAACCTCTCTGGAACAGCAGCCGACAGACAG	1246
QY	1138	GAGCGGTGTGCTCAGCTGAGACGCGCCGACAGAGAGAAAGACCGGAGACGCGACAGAG	1197
Db	1247	GAGCGCTGTGCGCAGCTGTGAGCGGGCGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAG	1306
QY	1198	CAGAGGCTCAAGCGGCACTGTGAGCTGTGAGAGCACTGTGAGAGACAGCGGAGCTGTGAG	1257
Db	1307	CAAGAGGCTCAAGCACTGTGAGCTGTGAGAGCACTGTGAGAGAGAGCGGAGCTGTGAG	1366
QY	1258	CGGCAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	1317
Db	1367	CGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	1426
QY	1318	CTGGAAGGACGACCACTTTGAATGTGGAAACGGAACCGGAGACAGGAATCTCTGAATCAG	1377
Db	1427	CTTGAAGGACCAACCACTTTGAATGTGGAAACGGAATGTGAAGGACAGGAATCTTAATCA	1486
QY	1378	AGGAACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	1437
Db	1487	AGGAACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	1546
QY	1438	GAGTTAGAGCTCTGAATGACAAAGACATCAGCTTAAGAGAGAGAGAGAGAGAGAGAG	1497
Db	1547	GAAATTAGAGCTCTTAATGTATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	1606
QY	1498	TGTGTGACTGTGCACTCCAGAGGCAAGAAATTTAGAGAGAGAGAGAGAGAGAGAGAG	1557
Db	1607	TGTGTGATTGTGACCCCAAGGCAAGAAATTTAGAGAGAGAGAGAGAGAGAGAGAGAG	1666

Oy 1558 ATGTGAAATCAACCCATTACAGCAGAGTTGACGAAATCTCAGAAATGCTTGAAGA 1617
 Db 1667 ATTGCCAAATACCCATCTACAGCAAAATTACGGAATCTCAGCAATGCTTGAAGA 1726
 Oy 1618 CTTATTTCAAGAAAACAGATACTCAGTGAACAGTTAAAACAAGTCCAGCAGAACAGTTTG 1677
 Db 1727 CTTATTTCAAGAAAACAGATACTCAGTGAACCAATTTAAAACAAGTCCAGCAGAACAGTTTG 1786
 Oy 1678 CATNAGACTCGCTCTTAACCTCAAAAGAGCTTGAAGCAAAAGAGTGGCCCGGAG 1737
 Db 1787 CACAAAGATTCACTTGTACCTTAAGAGCTTGAAGCAAAAGAGTGGCCCGGAG 1846
 Oy 1738 CAGCTCCGAGACAGCTGACAGAGTGAAGAGACCAAGTCAAGCTCAGAGAGATT 1797
 Db 1847 CACCTACAGACCACTGATGAATGAAGAAAGAACTAAGTCAAACTACAGAGATT 1906
 Oy 1798 GATGTTTTCAACACACAGCTGAAGAACTGAGAGATACATAGCAAAACAGCACTCAG 1857
 Db 1907 GATATTTTCAATATACAGCTGAAGAACTAAGAGAAATACAAATTAACCAACTCCAG 1966
 Oy 1858 AAGCAGAGTCCCTGAGAGCAGCGGACTGAAGCAAGAAAGACAGAGAGAGCCTG 1917
 Db 1967 AAGCAAAAGTCCATGAGAGCTGAACGACTGAACAGAAAGAAAGCAAGAAAGATCAT 2026
 Oy 1918 GAGTTAGAGAAAGCAAAAGGAAGCGCTCAGAGACGAGTTCAAGAAAGGAACAAGCAATGG 1977
 Db 2027 GAATTTGAAAACCAAAAAGAAAGAACCCAAAGACAGCTCAGAGAAAGGAACAAGAGTGG 2086
 Oy 1978 CTGAGCATGTGACAGAGAG---GAGCAGCAGCGCCCGGAAACCCCAAGAGAGAGC 2034
 Db 2087 CTGAGCATGTGACAGAGAGAGCAGAGCATCAGAGACCAAGAAACCTCCACAAAGAGAA 2146
 Oy 2035 AGACGTGAAGGGAAGACAGTGTCAAGAAAGAGAGGCGGAAGAGAGAGCCAGCGCGAA 2094
 Db 2147 AAATCTGAAAAGGAGAGAGTGTCAAAAAGAAAGATGCGAGAGAAAGGCAACAGAA 2206
 Oy 2095 ATGCAAGACAAGCAGAGTCCGCTTTTCATCCGATCAGAGCCAGCTAAGCTGCGCAC 2154
 Db 2207 GCACAAAGCAAGCTGGTGGCTTTTTCATCAACACAAAGAACAGCTAAGCAGCTGTC 2266
 Oy 2155 CAGGACCCCTGCTCAACACAGAGAAAGCGCTTACCATTTCTGCAACAGAGAGTGA 2214
 Db 2267 CAGGACCCCTGCTCAACACAGAGAAAGGTCCTTACCATTTCTGCAACAGAGAAATGA 2326
 Oy 2215 AAAGGTATATATTAACGAGCGCTGAACCCCTTTGAATCCAGAGACAGATGAATCAAC 2274
 Db 2327 AAAGGTGTATTAACCGGCACTGTACCTTTGAATCCAGAGACAGATGAATCAACT 2386
 Oy 2275 ATCCAGCCAGAGATATAGTCAT-----GTTGATGAAAGCCAGACTGGA 2319
 Db 2387 ATCCAGCCAGAGACATAGTCAATGTAAAGGGGAATGGTGGATGAAAGCCAAACTGGA 2446
 Oy 2330 GAGCAGATGGCTTTGAGAGAGAGCTGAAGGGAAGACGGAATGTTCCCTGCAAACTAT 2379
 Db 2447 GAACCCGCTGCTTGAAGAAATTAAGAAAGACAGGTTGTTCCCTGCAAACTAT 2506
 Oy 2380 GAGAGAAAGATTCCAGAAATGAGGTCCCACTCCAGCCAAACAGTGAACGATCTGCA 2439
 Db 2507 GAGAGAAATATCCAGAAATGAGGTCCCGCTCCAGGTGAACAGTGAATCTGATCAACA 2566
 Oy 2440 TCTGCCCCCTGCCCCAACTGGCTGTGCTGAGACCCCTGCTCTTTGCAAGTGAACCTT 2499
 Db 2567 TCTGCCCCCTGCCCCAACTGGCTGTGCTGAGACCCCGCCCTTTTGGCAGTAACTCT 2626
 Oy 2500 TCTGAGCCCTTCACAACCCCAACAACCTGGAGACCTTCAAGTTCAACGTTGCGCCAGAC 2559
 Db 2627 TGAAGCCCTTCACAACCCCAACAACCTGGAGACCTTCAACGTTGCGCCAGAC 2686
 Oy 2560 TCAAAACGAGAGACCAAGAAACGACCACTGGAGTAAGTGGCGGCTCAGCTTCTCTGACC 2619
 Db 2687 AGGAATGAGAAACAGAAACGATATCTGGAGTGAATGGGAGCCAGCCCTCTCTTCAAC 2746
 Oy 2620 GTACTAGTGTGCGCACTTACGAGCAGATCAGCCTTTACCCAGCAACAGCACTGAC 2679

Db 2747 GTTCAAAGTCCCGGCACTTAAGCAGAGGTCGCTTTTCTCAGCAACGCGCACTGGC 2806
 Oy 2680 TCTTCCCATATCTCCCTGCTGGGCTGAGGTAAGAAAGTGAAGGCTCAACGCAAGCC 2739
 Db 2807 TCTTCCCGCTCTCTGTGCTAGGCTGAGGTAAGAAAGTGAAGGCTCAACGCTCAAGCC 2866
 Oy 2740 CTGTATCCCTGAGAGGCAAAAAGACAAACCACTTAATTTTAAACAAAGTGAAGTATC 2799
 Db 2867 CTATATCTTTGAGAGGCAAAAAGACAAACCACTTAATTTTAAACAAAGTGAAGTATC 2926
 Oy 2800 ACCGTTGGAACAGCAAGACATGTGTGTTTGAAGAGTTCAAGTCAAGAGGTTTG 2859
 Db 2927 ACGTCTGGAACAGCAAGACATGTGTGTTTGAAGAGTTCAAGTCAAGAGGTTTG 2986
 Oy 2860 TTCCCAAGTCTTACGTGAATCTATTTCAAGGCGCCGTAAAGAAATCAACAGTATGAT 2919
 Db 2987 TTCCCAAGTCTTACGTGAATCTATTTCAAGGCGCCGTAAAGAGTCAACAGATGAT 3046
 Oy 2920 ACTGCGCTTCTGAAGTCCGCTAGTCTAAGAGAGTGGCTTCCCGGCGCCCAAGCA 2979
 Db 3047 TCTGTTCTTCAAGAGTCTTCTAGTCTTAAGGAGTACCTTCCAGACCCCAAGCC 3106
 Oy 2980 GCCATTCGCGAGAGAGTTTATTTGCCATGTACAATACAGAGTTCTGACAGAGAT 3039
 Db 3107 GTGTTTCGGAAGAAATTTATTTGCCATGTACAATACAGAGTTCTGACAGAGAT 3166
 Oy 3040 TTAACTTTTACAGAGGAGATGTATTTGTGTTTACAAAGAAAGTGTGACTGTGACG 3099
 Db 3167 TTAACTTTTACAGAGGAGATGTATTTGTGTTTACAAAGAAAGTGTGACTGTGACG 3226
 Oy 3100 GGAACGGTGGGCGCAAGTCCGAGTCTTCCCTTCAATATGAGAGCTTTAAAGATTCA 3159
 Db 3227 GGAACAGTGGGCGCAAGGCGGAGTCTTCCCTTCAATATGAGAGCTTTAAAGATTCA 3286
 Oy 3160 GAGGCTCTGGAACCTGTGGGAAAACAGGAGTTTAAAGAAAACCTGAATTTGCCAG 3219
 Db 3287 GAGGCTCTGGAACCTGTGGGAAAACAGGAGTTTAAAGAAAACCTGAATTTGCCAG 3346
 Oy 3220 GTTATTTGCTTCTACGCTGTGCTACGTGTCCTGGAACATCACTCTGCTCTGCGCAGCTG 3279
 Db 3347 GTTATTTGCTTCTACGCTGTGCTACGCTGCGCAGCTGCTCTGCTCTGCGCAGCTG 3406
 Oy 3280 ATTCTATCCGAGAAAAGAAACCCAGTGAATGTTGGAAGAGAACTGCAAGCTCGAGGG 3339
 Db 3407 ATTCTATCCGAGAAAAGAAACCCAGTGAATGTTGTAAGTTAAAGCTTTAAACCTGAGAG 3466
 Oy 3340 AAAAAGCGCAGATAGGTGTTTCCAGCAATTAATGTTCAACTTTAGCCTCCGAGAA 3399
 Db 3467 AAAAAGCGCAGATAGGTGTTTCCAGCTAATTAATGTAAGCTTTAAACCTGAGAG 3526
 Oy 3400 AGCAAAATCAACCCCAACCTAGACTACCAAGACCGAGTGCAGCAGAGAGTGTGCAAGTG 3459
 Db 3527 AGCAAAATCACTCCCAAGAGCCACTTAAGTCAACACACTTAAGGAGAGTGTGCAAGTG 3586
 Oy 3460 ATCGAGATGATCACTACCGCCAGAGAGATGACGAACTAGGCTTCAAGAAAGGCGAG 3519
 Db 3587 ATTGGAATGATCACTACCGCCAGAGATGATGATGATGATGATGATGATGATGATGATG 3646
 Oy 3520 ATCATCAACGCTCTCAACCAAGAGAGACCGGACTGTGGAAGAGAGAGTCACTGGCA 3579
 Db 3647 ATCATCAACGCTCTCAACCAAGAGAGACCGGACTGTGGAAGAGAGTCACTGGCA 3706
 Oy 3580 GTTGGGCTTTCCATCAATTAATGTAAGTGTGACCAAGACATGAGACCCAGCAGCA 3639
 Db 3707 GTGGGCTTTCCATCAATTAATGTAAGTGTGACCAAGACATGAGACCCAGCAGCA 3766
 Oy 3640 TGA 3642
 Db 3767 TGA 3769

RESULT 7

AF064244 7247 bp mRNA linear PRI 21-NOV-1998
 LOCUS Homo sapiens intersecin long form mRNA, complete cds.
 DEFINITION AF064244
 ACCESSION AF064244
 VERSION AF064244.1 GI:3859854
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 7247)
 AUTHORS Guipponi, M., Scott, H. S., Chen, H., Schebesta, A., Rossier, C. and Antonarakis, S. E.
 TITLE Two isoforms of a human intersecin (ITSN) protein are produced by brain-specific alternative splicing in a stop codon
 JOURNAL Genomics 53 (3), 369-376 (1998)
 MEDLINE 99017974
 PUBMED 9799604
 REFERENCES 2 (bases 1 to 7247)
 AUTHORS Guipponi, M., Scott, H. S., Chen, H., Schebesta, A., Rossier, C. and Antonarakis, S. E.
 TITLE Direct Submission
 JOURNAL Submitted (05-MAY-1998) Genetics and Microbiology, CMU, 1 rue Michel-Servet, Geneva 4 CH-1211, Switzerland
 FEATURES
 source
 1. 7247
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /chromosome="21"
 /map="21q22.1-q22.2"
 /cissue_type="brain"
 /dev_stage="fetus"
 107..5272
 /codon_start=1
 /product="intersecin long form"
 /protein_id="AAC78611.1"
 /db_xref="GI:3859854"
 /translation="MAOPPTREGSLDIAIVTEERAKHDOFHSLKISPTITGQA
 RNPFPSGIPQVLAQIWMALAMNNDGRDQVFSIAMKLIKLOGLVPSLPVPM
 KOQPAVISAAPFGMGJASMPPLTAVAPMGSIIPVMSPTLVSSVTPAVPLAN
 GAPVLOPLPAPFAPPAATLPSKSSFSRSGSGLNTLOKASFDVASVPVAMVAP
 OSSRLKROLFNHSDKMSGHLTGPOARTILMOSLPOKOLASINLSIDIDOGKULTA
 EEPILAMHLIDVMSGPLPVLPPEVTPSPRRVRSVSGSIVISSTVDORLPEPV
 LEDBOOULKPLPVTFEDKRENFERNIELEKROALLBQKQERLAQBRACQ
 KREBERQBERKQLELEKQERQERERERKEIEREARELERERQLEME
 RRRQELNQNKQEDIVYLKAKKTLFEELANDKHGLEKLDIRCLITTORO
 EISTKSRLELAETIHLQOOLQESQOMGLIPKQIILNDQLKVOQNSLHRLSLV
 TLKRALKELELAHROHLRDOLDEVERETRSKLOEIDIPNNOLKELEIHNKQOLKOKS
 MEAERLKQKQERKITELEKQEREAORROERDQWLEHVQOEDERORPKLIEEKL
 KVEESYKQDGEKQERQODKGLRFRHOEPKAPAVAPNSTAEKQPLTISAGENV
 KAVYTRALYPFESRSHDEITIQGDIIVNKGWVDSQGEPMGLKNGKMPA
 NNAEKIPENEVAPVAVPVDSTSAAPAKLARETPALVTSSEPTNNNADFSST
 WPTSTKEPETNDWMAAOPSLTVPASQLOKRSAPATATASPSVYLQGEKVE
 GLOAOLYPMRAKQNLNKNKDVITVLEODMNGEVOGKGFPSYKVLISGP
 IRTKSTMDGSSSPASLKRVAAPKPVASGEPIAMTVTSSGSEKLTPOGQDVL
 VTKKDDMMTGTGDKAGVPSPSYVALKQSBGSGTAKGSLKREKELIQAIVASTAT
 GPEQLTAPGQLLIRKNKPGWESGLDARGKQKQIGMPAYVALNLPNGTSKTPP
 EEPKSTALAAVOCVIMDYTAQNDDELAFNKGQIINVANKEDPMWKEVNGVGLF
 PSNYVLTITDMDPSQOMCSDLHLDMITPERKQGYIHLLIVTEBNVNDLQVLEI
 FOKPLMESELDTKEVAMI FVNKELIMCNILKALRVKQKSGKMPVKNIGDILS
 AOLPHMOPIYRFGSROLNGAALIQCKTDAPEKFEVRLKLEMDPRCKMPISLRLKP
 MORVTRYPILIKRILENTBNHDSHLKALEKAEKLSQVNEGVREKENSRLKEMT
 QAHVQEGESLEQVFNSTNCGPRKFLSHGKLYKAKNKELYGFENDPLLTQITK
 PLGSGTDVFPSPKSNLYKMYKTPILFLEBVVLKPLDPSGDEPIFIHSHIRVYLR
 AESINERTAMVQKIKAASELILETEKKEKAVLRSQATGIGRAIVANVAVGIEKP
 CSNHSNRYCEVTGMSOCHITKTIQDITNPKNNSCOPIRDLBEQVCLTVEPDDO
 FSPDDFLGRTEIRVADIKKQSGSKGVTKCLLHVEPTGEIYVRDLQLEPDR"
 misc_feature
 167..406
 /note="encodes EH domain"
 misc_feature
 767..1936
 /note="encodes EH domain"
 misc_feature
 2324..2524

/note="encodes SH3 domain"
 2843..3019
 /note="encodes SH3 domain"
 3110..3286
 /note="encodes SH3 domain"
 3326..3520
 /note="encodes SH3 domain"
 3569..3748
 /note="encodes SH3 domain"
 3836..4390
 /note="encodes SH3 domain"
 4649..4819
 /note="encodes SH3 domain"
 4895..5143
 /note="encodes SH3 domain"
 1652 c 1747 g 1692 t
 BASE COUNT 2156 a 1652 c 1747 g 1692 t
 ORIGIN
 Query Match 77.6%; Score 2826.8; DB 9; Length 7247;
 Best Local Similarity 86.4%; Pred No. 0;
 Matches 3164; Conservative 0; Mismatches 477; Indels 21; Gaps 3;
 1 ATGGCTCAGTTCCACACCTTCGCTGAGCTGATGTCGGCCATTAAGTGAG 60
 107 ATGGCTCAGTTCCACACCTTCGCTGAGCTGATGTCGGCCATTAAGTGAG 166
 61 GAAAGGCCAAGCATGACGACAGTTCCTTACCTGAGCCATGACGGATTTATT 120
 167 GAAAGGCCAAGCATGATGACGACAGTTCCTTACCTGAGCCATTAAGTGAG 226
 121 GGTGATCAGCAGGAACTTTTTCATCTGGGTTACCTCAGCCTGCTTAGACAA 180
 227 GGTGATCAGCAGGAACTTTTTCATCTGGGTTACCTCAGCCTGCTTAGACAA 286
 181 ATATGGGCGCTAGCGGACATGATTAAGATGAGAGATGATCAATGATTTTCCATA 240
 287 ATATGGGCGCTAGCGGACATGATTAAGATGAGAGATGATCAATGATTTTCCATA 346
 241 GCATGAGAGCTTATCAACTGAGATCAAGATATCAGCTCCCTCCACACTTCCCT 300
 347 GCATGAGAGCTTATCAACTGAGATCAAGATATCAGCTCCCTCCACACTTCCCT 406
 301 GCATGAGAGCTTATCAACTGAGATCAAGATATCAGCTCCCTCCACACTTCCCT 360
 407 GCATGAGAGCTTATCAACTGAGATCAAGATATCAGCTCCCTCCACACTTCCCT 466
 361 GGTAGATGCGACACCTGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
 467 GGTAGATGCGACACCTGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 526
 421 GGAATGTCCTCAACCTTATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 480
 527 GGAATGTCCTCAACCTTATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 586
 481 GGGGGCTCCCGGCTATGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
 587 GGGGGCTCCCGGCTATGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 646
 541 AAGAGTTCCTTCTGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600
 647 AAGAGTTCCTTCTGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 706
 707 GCAAGTTCCTTCTGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 766
 601 GCAAGTTCCTTCTGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
 707 GCAAGTTCCTTCTGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 766
 661 TCAAGGCTGAATTAAGAGAGTATTAAGAGAGAGAGAGAGAGAGAGAGAGAG 720
 767 TCAAGGCTGAATTAAGAGAGTATTAAGAGAGAGAGAGAGAGAGAGAGAGAG 826
 721 ACAAGTCCCGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 780
 827 ACAAGTCCCGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 886

781 TCAATATGAAATCTTTTGAACATTTGATGAAGATGGAATACTCATCTGCAAGAAATTTATC 840
 Db TCAATATGAAATCTTTTGAACATTTGATGAAGATGGAATACTCATCTGCAAGAAATTTATC 946
 841 CTAGCTATGCACTTAATGATGTTGTCATGTCGTGTCAGCCACTGCCCCCTGCTGCT 900
 Db CTGGAATGCACTTATGATGTTGTCATGTCGTGTCAGCCACTGCCCCCTGCTGCT 1006
 901 CCAGATATCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 960
 Db CCAGATATCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1066
 961 AGCTTTCTCTCTGATGATGAGGCTGCTGAGGCTGCTGAGGCTGCTGAGGATGAGGAG 1020
 Db AGCTTTCTCTCTGATGATGAGGCTGCTGAGGCTGCTGAGGCTGCTGAGGATGAGGAG 1126
 1067 AGCTTTCTCTCTGATGATGAGGCTGCTGAGGCTGCTGAGGCTGCTGAGGATGAGGAG 1126
 1021 C---CAGAGAGAACTGCTGCTGATGATTTGAAAGTAAAGAGCGGAGAACTTCAGGCA 1077
 Db CAATTAGAAAGAAATTAACCTGTAACGTTTGAAGTAAAGAGCGGAGAACTTCAGGAGT 1186
 1078 GGCAGTGTGAGCTGAGAAAGCGCGCAAGCGCTCTTTGAGCAGCAGCGCAAGAGAG 1137
 Db GGCAGTGTGAGCTGAGAAAGCGCGCAAGCGCTCTTTGAGCAGCAGCGCAAGAGAG 1246
 1187 GGCAGTGTGAGCTGAGAAAGCGCGCAAGCGCTCTTTGAGCAGCAGCGCAAGAGAG 1246
 1138 GAGCGGTGCTGCTGAGTGAAGCGCGCGCAAGCGAGAGAGAGAGAGAGAGAGAGAG 1197
 Db GAGCGGTGCTGCTGAGTGAAGCGCGCGCAAGCGAGAGAGAGAGAGAGAGAGAGAG 1306
 1247 GAGCGGTGCTGCTGAGTGAAGCGCGCGCAAGCGAGAGAGAGAGAGAGAGAGAGAG 1306
 1198 CAGAGGCGCAAGCGCGAGCTGAGAGCTGAGAGAGCTGAGAGAGAGAGAGAGAGAG 1257
 Db CAAGAGCGCGCAAGCGAGCTGAGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAG 1366
 1307 CAAGAGCGCGCAAGCGAGCTGAGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAG 1366
 1258 CGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1317
 Db CGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1426
 1367 CGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1426
 1318 CTGGAAGGCGCGCAAGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1377
 Db CTGGAAGGCGCGCAAGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1486
 1427 CTGGAAGGCGCGCAAGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1486
 1378 AGGAACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1437
 Db AGGAACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1546
 1487 AGGAACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1546
 1438 GAGTTAGAGAGCTTGAATGACAAAAGCATCAGTTAGAGAGAGAGAGAGAGAGAGAG 1497
 Db GAGTTAGAGAGCTTGAATGACAAAAGCATCAGTTAGAGAGAGAGAGAGAGAGAGAG 1606
 1547 GAGTTAGAGAGCTTGAATGACAAAAGCATCAGTTAGAGAGAGAGAGAGAGAGAGAG 1606
 1498 TGTGCACTGAGCAAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1557
 Db TGTGCACTGAGCAAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1666
 1607 TGTGCACTGAGCAAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1666
 1558 ATTGCTGAAGTCAACCACTTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1617
 Db ATTGCTGAAGTCAACCACTTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1726
 1667 ATTGCTGAAGTCAACCACTTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1726
 1618 CTTATTTCCAGAGAAACAGATCTCAGTGAACAGTTAAACAGGTCAGAGAGAGAGAG 1677
 Db CTTATTTCCAGAGAAACAGATCTCAGTGAACAGTTAAACAGGTCAGAGAGAGAGAG 1786
 1727 CTTATTTCCAGAGAAACAGATCTCAGTGAACAGTTAAACAGGTCAGAGAGAGAGAG 1786
 1678 CATAGAGAGCTGCTCTTAACCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1737
 Db CATAGAGAGCTGCTCTTAACCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1846
 1787 CATAGAGAGCTGCTCTTAACCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1846
 1738 CAGCTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1797
 Db CAGCTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1906
 1847 CAGCTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1906
 1798 GATGTTTTCACACACAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1857
 Db GATGTTTTCACACACAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1966
 1907 GATGTTTTCACACACAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1966

1858 AAGCAGAGTCCCTGAGAGCGCGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGCTG 1917
 Db AAGCAGAGTCCCTGAGAGCGCGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGCTG 2026
 1918 GAGTTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1977
 Db GAGTTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2086
 2027 GATTTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2086
 1978 CTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2034
 Db CTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2146
 2087 CTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2146
 2035 AGACTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2094
 Db AGACTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2206
 2147 AACTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2206
 2095 ATGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2154
 Db ATGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2266
 2207 GCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2266
 2155 CAGGACCTCTGCTTACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2214
 Db CAGGACCTCTGCTTACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2336
 2267 CAGGACCTCTGCTTACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2336
 2215 AAGTGTATATTACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2274
 Db AAGTGTATATTACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2386
 2327 AAGTGTATATTACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2386
 2275 ATCCAGCCAGAGAGATATAGTAT-----GCTGATGAAGAGAGAGAGAGAGAGAG 2319
 Db ATCCAGCCAGAGAGATATAGTAT-----GCTGATGAAGAGAGAGAGAGAGAGAG 2446
 2287 ATCCAGCCAGAGAGATATAGTAT-----GCTGATGAAGAGAGAGAGAGAGAGAG 2446
 2220 GAGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2379
 Db GAGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2506
 2447 GAGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2506
 2380 GAGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2439
 Db GAGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2566
 2507 GAGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2566
 2440 TCTGCCCCCTGCCCCCAAGCTGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2499
 Db TCTGCCCCCTGCCCCCAAGCTGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2626
 2567 TCTGCCCCCTGCCCCCAAGCTGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2626
 2500 TCTGAGCCCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2559
 Db TCTGAGCCCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2686
 2627 TCAAGAGCCCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2686
 2560 TCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2619
 Db TCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2746
 2687 TCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2746
 2620 GTACCTAGTGTGCGCAGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2679
 Db GTACCTAGTGTGCGCAGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2806
 2747 GTACCTAGTGTGCGCAGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2806
 2680 TCTGCCCCCTGCCCCCAAGCTGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2739
 Db TCTGCCCCCTGCCCCCAAGCTGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2866
 2807 TCTGCCCCCTGCCCCCAAGCTGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2866
 2740 CTGATCCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2799
 Db CTGATCCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2926
 2887 CTGATCCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2926
 2800 ACCGTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2859
 Db ACCGTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2986
 2927 ACCGTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2986
 2860 TTCCCCCAAGTTTACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2919
 Db TTCCCCCAAGTTTACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3046
 2987 TTCCCCCAAGTTTACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3046
 2920 ACTGAGCCCTTACTGAAGTCTGCTAGTCTAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2979

QY	421	GGAAATGTCTCCACCCTTATGTATCTTCTGTCCCTCCAGACGAGATGTCCTCCCTGGCTATAC	480
Db	461	GGAAATGTCTCCGCCCTTATGTATCTTCTGTCCCTCAAGAGAGATGTCCTCCCTGGCTATAC	520
QY	481	GGGGGTCTCTCCCGGTATACAGGCTCTGTGCTGGATTGGGCATTCCTGACGCCACATGGCCA	540
Db	521	GGGGGCCCTCTGTATATACAGCTCTGTGCTGATTTGCTCATCTCTGACGCCACATTTGCCA	580
QY	541	AAGAGTTCTTCTTTCAGCAGATCTTGCTCAGGGGTACAAATTAAACACTTAAGTTACAGAA	600
Db	581	AAGAGTTCTTCTTTCAGCAGATCTTGCTCAGGGGTACAAATTAAACACTTAACAGAA	640
QY	601	GCACATATATTCGATGTGTGCCAGCGCCCTCCAGACGAGAAATGGGGCTGTGCTTCAGTCA	660
Db	641	GCACATATATTTGAATGTATGCCAGCGCCCTCCAGACGAGAAATGGGGCTGTGCTTCAGTCC	700
QY	661	TCAAGGCTGAAATATACAGGCAAGTTATTTCAACAGCCACGACAAATCTATAGTGGACACTTA	720
Db	701	TCAGAGCTGAAATATACAGGCAAGTTATTTCAACAGCTACGACAAATCATATAGTGGACACTTA	760
QY	721	AACAGGTCCCAAGGCAAGAACTATTTCTCATGCAATCAAGTTTAAACCCAGAGCTCAGCTGGCT	780
Db	761	AACAGGTCCCAAGGCAAGAACTATTTCTCATGCAATCAAGTTTAAACCCAGAGCTCAGCTGGCT	820
QY	781	TCGAATATGGAATCTTTTCTGCACTTTGATCAAGATGGAAAACTCAGCTGCAAGAAATTTATC	840
Db	821	TCGAATCTGGAATCTTTTCCGACATTTGATCAAGATGGAAAACTCACCGCAAGAAATTTATC	880
QY	841	CTAGCTATGACCTTAATTTGATTTGTCGCAATGTGTGGTCAAGCCATGCGCCCGTCCGTGCT	900
Db	881	CTAGCGATGACCTTATTTGATTTGTCGCAATGTGTGGTCAAGCCATGCGCCCGTCCGTGCT	940
QY	901	CCAGAAATCATCCCTCTCTTCTTCCAGAAAGATTGCTCCCGCAGTGGAGTGTCCGTACTA	960
Db	941	CCAGAAATCATCCCTCTCTTCTTCCAGAAAGATTGCTCCCGCAGTGGAGTGTCCGTACTA	1000
QY	961	AGCTTCTTCTTGTGTGATCAAGGCTGCTGCTGAGAGCCGTGTCAAGAGATGACGACAG	1020
Db	1001	AGCTTCTTCTTGTGTGATCAAGGCTGCTGCTGAGAGCCGTGTCAAGAGATGACGACAG	1060
QY	1021	CCAGAGAAAGAAATCTGCTGTGACATTTTAAGATTAAGAACGGGGAGAACTTGTGAGCGAGGC	1080
Db	1061	GTCGAAAAAGAAAGCTCTGTCGTGACATTTTAAGATTAAGAACGGGGAGAACTTGTGAGCGAGGC	1120
QY	1081	AGTGTGAGCTGGAGAAAGCGCGCCAGCGGCTCTTGGACGACGCGCAAAAGACGACGAG	1140
Db	1121	AACCTGTGAGCTGGAGAAAGCGCGGAGCGAGCGCTCTCTGTGAGCAGCAGCGCAAGAGCAGAG	1180
QY	1141	CGGTTGTGCTCAGCTGTGAGGCGCGCCGACGACGAGAGAGAGAAAGCGGGAACGCCAGGACAG	1200
Db	1181	CGGTTGTGCTCAGCTGTGAGGCGCGCCGACGAGAGAGAGAGAAAGCGGGAACGCCAGGACAG	1240
QY	1201	GAGGCGCAAGCGGCACTGTGAGGTTGGAGAAAGCACTGTGGAAAGCAACGGGAGGCTGTGAGCGG	1260
Db	1241	GAGGCGCAAGAGCACTGTGAGGTTGGAGAAAGCACTGTGGAAAGCAACGGGAGGCTGTGAGCGG	1300
QY	1261	CAGCGAGAGAGAGAGAGAGAGAGATTCGAGAGCGCGCGAGGCGCGCAAAACGGGAACTTG	1320
Db	1301	CAGAGAGAGAGAGAGAGAGAGAGATTCAGAGCGGCGGAGCGGCGCAAAACGGGAGGCTTC	1360
QY	1321	GAAAGCAGCGCACTTGAATGTGGAAACGGAAACGGGAGCAGGAATCTCTGAATCAGAGG	1380
Db	1361	GAAAGCAGCGCACTTGAATGTGGAAACGGAAACGGGAGCAGGAATCTCTGAATCAGAGG	1420
QY	1381	AACAAAGAGAGAGAGAGGCGACCTGTGCTTGAAGGCAAGAGAGABAATCTGTGAGTTTGA	1440
Db	1421	AACAAAGAGAGAGAGAGGCGCATCTGTGCTTGAAGGCGAGAGAGAACTCTGTGAGTTTGA	1480
QY	1441	TTAGAGAGCTCTGAATGACAAAAAGATCAGCTGAAGAGAAAACTTCAGAGATTCACGAGT	1500
Db	1481	TTAGAGAGCTCTGAATGACAAAAAGATCAGCTGTGAAGAGAAAACTTCAGAGATTCACGAGT	1540
QY	1501	CGACTGTGCAACCCAGAGGCAAGAAATTGAGACAGCAAGAACAGTCTAGAGAGCTAAGATTT	1560

Dp	1541	CGACTGGCAACCCAGAGCGAAGAAATTGAGCGACAAACAAATCTTAGAGCTGAGAAATT	1600
Qy	1561	GCTGAAATCAACCCACTTACAGCAGCAGTGTGCAAGATCTCAGCAAAATGCTTGGAAAGCTT	1620
Dp	1601	GCGGAATCAACCCACTTAGACCAACAGTTGACGGAATCTCAGCAGATGCTTGGAAAGCTT	1660
Qy	1621	ATTCCAGAAACAGATCTCAGTGACCAAGTTAAACAAGTCCAGCGAACAATTTCAT	1680
Dp	1661	ATTCCAGAAACAGATCTCAGTGACCAAGTTAAACAAGTCCAGCGAACAAGTTTCAT	1720
Qy	1681	AGAGACTGCTTCTTACCCTCAAAAGCCTTGGAAAGCAAGAGAGCTGGCCCGGACAG	1740
Dp	1721	AGAGATTGCTTCTTACCCTCAAAAGCCTTGGAAAGCAAGAGAGCTGGCCCGGACAG	1780
Qy	1741	CTCCGGAGCAGCTGGAAGAGGTGAGAGAGACCAAGTCAAGCTTCAGAGATTTGAT	1800
Dp	1781	CTTGAGAGAGCAGCTGGAAGAGGTGAGAGAGACCAAGTCAAGCTTCAGAGATTTGAT	1840
Qy	1801	GTTTTCACAAACAGCTGGAAGAACTGAGAGATACATAGCAAAACAGCACTCCAGAG	1860
Dp	1841	GTTTTCACAAACAGCTGGAAGAACTGAGAGATACATAGTAACAGCAGCTCCAGAG	1900
Qy	1861	CAGAGTCCCTGGAGGCGAGCGCGACTGAAGCAGAAAGCGAGAGAGGAAAGCTGGAG	1920
Dp	1901	CAGAGTCCCTGGAGGCGAGCGCGACTGAAGCAGAAAGCGAGAGAGGAAAGCTGGAG	1960
Qy	1921	TTAGAGAAAGCAAAAGSAAAGCGCTCAGAGACAGATTCAGAAAGSAAAGCAATGCGTG	1980
Dp	1961	TTGAGAGAAAGCAAAAGSAAAGCGCTCAGAGACAGATTCAGAAAGSAAAGCAATGCGAG	2020
Qy	1981	GAGCATGTGCAGCAGAGAGAGCAGCCACGCCCGGAAACCCACAGAGAGACAGACTG	2040
Dp	2021	GAGCATGTGCAGCAGAGAGAGCAGCCACGCCCGGAAACCCACAGAGAGACAGACTG	2080
Qy	2041	AAGAGGAAAGCAGTGTCAAGAAAGAGGCGGAAAGAGAGAGCAGACCGGAAATGCA	2100
Dp	2081	AAAAGGAAAGCAGTGTCAAGAAAGAGGCGGAAAGAGAGAGCAGACCGGAAATGCA	2140
Qy	2101	GACAAGCAGATCGGCTTTTCCATCCGATCAGAGCAGCTAAGCTTGCCACCCAGCA	2160
Dp	2141	GACAAGCAGATCGGCTTTTCCATCCATCCATCAAGACCAAGCTTAAGCCG---CCGAGCA	2197
Qy	2161	CCCTGCTTACACAGAGAAAGGCCCGCTTACATTTCTGCACAGAGAGTGTAAAGTG	2220
Dp	2198	CCCTGCGCCACAGACAGAAAGAGTCCGCTTCAATCTTCGACAGAGAGTGCCAAAGTG	2257
Qy	2221	GTAATTTACCGAGGCGCTGACCCCTTTGAATCCGAAGTCAAGTAGATTAACATCCAG	2280
Dp	2258	GTAATTTACCGAGGCGCTGACCCCTTTGAATCCGAAGTCAAGTAGATTAACATCCAG	2317
Qy	2281	CCAGAGATTAATGATCAT-----GGTGAATGAAGCCAGACTGAGAGGCCA	2325
Dp	2318	CCAGAGATTAATGATCATGTTAAAGGGGAATGGGTGATGAAGCCAGACCGGAGAGCCA	2377
Qy	2336	GGATGCTTGGAGAGAGCTGAAAAGGAGAGCGGATGCTTCCCTGCACAACTATATGCAGAA	2385
Dp	2378	GGATGCTTGGAGAGAGCCGAAAGGAGAGCAGAGATGCTTCCCTGCACAACTATATGCAGAG	2437
Qy	2386	AAAGTTCCAGAAATGAGGTTCCCACTCAGCCCAAACAGAGACCGATCTGACATCTGCC	2445
Dp	2438	AAAGTTCCAGAAATGAGATTCACACTCCACCAAACAGAGACCGATCTGACATCTGCC	2497
Qy	2446	CCGCGCCCAAACTGGGCTCTGGGGAGAGCCCTGCTCTTGGCAGATGACCTTCTTGAG	2509
Dp	2498	CCGCGCCCAAACTGGGCTCTGGGGAGAGCCCTGCTCTTGGCAGATGACCTTCTTGAG	2557
Qy	2506	CCCTTCACAAACCCCAACAACTGGGAGACTTCAAGTTCAAGTGCCAGCAGCTCAAC	2565
Dp	2558	CCCTTCACAAACCCCAACAACTGGGAGACTTCAAGTTCAAGTGCCAGCAGCACAAT	2617
Qy	2566	GAGAAAGCAGAAACGACCACTGGGATTCGTGGCGGCTCAGCCTTCTTGACGATACCT	2628


```

misc_feature      3364..3558
                    /note="encodes SH3D domain"
misc_feature      3607..3786
                    /note="encodes SH3E domain"
BASE COUNT       1290 a 920 c 1029 g 856 t 8 others
ORIGIN
Query Match      55.7%; Score 2027.6; DB 5; Length 4103;
Best Local Similarity 73.6%; Pred. No. 0;
Matches 2685; Conservative 4; Mismatches 918; Indels 39; Gaps 7;

QY      1 ATGGCTCAGTTTCCACACCTTTCGGTGTAGCCCTGATGTCTGGGCCATTACTGTGAG 60
DB      193 ATGGCTCAGTTTGGAACTCCGTTGGGGGTATTGACATCTGGGCATTAACGCTGAG 252
QY      61 GAAAGGCCAAGCATGACGACGACGATTCCTTAGCCCTGAAGCCGATATGCGGATTTATCT 120
DB      253 GAAAGGCTTAAACATGACGACGACGATTCATGGGCTCAAGCCAAACAGCTGATATATCA 312
QY      121 GGTGATCAAGCGAGGAACTTTTTCATCTGGGTTACCTCAGCCTGTCTTAGCAAA 180
DB      313 GGTGACCAAGCTAGGAAATTTTTCCTTAGTCCGAGCTCCCGACCTGTGTGCAAG 372
QY      181 ATATGGGCGCTAGCGGACATGATTAACGATGGAAGATGATCAAGTGAATTTTCCATA 240
DB      373 ATATGGGCGCTGTGTCATGATGAACATGATGGAAGATGACCAACCTGAGTTCTCATA 432
QY      241 GCCATGAAGCTTATTAACCTGAAGCTACAGGATTAAGCTCCCTCCACACTTCCCT 300
DB      433 GCTATGAATTAATTAACCTTAACCTTAACAGTTAACCGTTGCCCTCTATTCTCCCTTCC 492
QY      301 GTCATGAACAGCAACAGTGGCTAT-----TTCCAGTGACCGACGATTTGTATAGA 354
DB      493 AATATGTGAAGCAACCGGTGTCATGCTGCTGTCTGTGTGGGGGTGGGATGAGT 552
QY      355 GGGATGTGTAGCATGACCACTCACTCACTGTGTGCTGTGTCCTGTCATGAGGCTTCA 414
DB      553 GGAATCGTGGGATTCCTCCCTAGCACTGATGAGCCCTGTGCAATCACTCATCCCA 612
QY      415 GTTGTGGAATGTCTCCACCTTATGATCTTGTGTCCTCCAGAGCAAGTGTCTCCCTG 474
DB      613 GTAGTAGGAATGTCTCTCTCTGTGTGTCTCTGTGTCCTCA-----CAGTTTCCACCTTGG 666
QY      475 GCTAACGGGGGCTCTCCCGTCATACAGGCTCTGCGTTTGGCATCTGACGCCA 534
DB      667 TCTAACGGGGGCTCTCGACGTATTAAGTCTCATCTGCTTTGCTCACTCT---GCCACA 723
QY      535 TGGCCAAAGAGTCTTCTTCTAGCAGATCTGTTCAGGGGTCAAACTTAACACTAAGTTA 594
DB      724 TTGCCAAAGAGTCTTCTTCTAGCGGCGCTCAGTTGCGGGTCCAGATTAACACTAATT 783
QY      595 CAGAAGGCACAATCATTTGATGTGCGCAGCGCCCTCCAGACAGCAAGTGGGCTGTGCT 654
DB      784 CAGAAAGCTCAGTCTTGTGATGTTCAGCCCCCCTCACTTGTGTAATGGGCTGTGCA 843
QY      655 CAGTCATCAAGGCTGAATTAACAGGAGTTATTCAACAGCAAGCAAAACTATAGTGA 714
DB      844 TCGTCTTAAAGATTGAAGTAAGACAAATTTTCAACAGCAGCAAAACCAATGATGAG 903
QY      715 CACTTAACAGGTCCCGACGCAAGAACTATTCTCATGCAATCAAGTTTACCCGAGCTCAG 774
DB      904 AATTAAACAGGCTCTCAGGCAAGAACTATTCTTATGAGCTCAAGTTTACCCGAAAGCAG 963
QY      775 CTGGCTTAATATGGAATCTTTCTGACATGTGATCAAGATGAAAACTCACTGCAAGAA 834
DB      964 CTAGCTTAACATATGGAATCTTTCAACATTAACCAAGATGAAAACTCAACGCCAAGAA 1023
QY      835 TTATCTAGCTATGACCTATATGATGTGTCATGTGTCAGCCGACCTGCGCGCTC 894
DB      1024 TTATATATAGCTATGACCTTATATGATGTGTCATGTGTCAGCCGACCTTCTCTCAATC 1083
QY      895 CTGCTCCAGATATCATCTCTCTCTCTTCAAGAAAGTTGCTCCGCGCAGTGGATGTCC 954

```

```

DB      1084 CTGCTCCAGAGTATATTTCCCATCTTTTGAAGAGTTCGATCAGGAGATGGTTATCC 1143
QY      955 GTCATAGCTCTTCTTCTGTGATGATCAGAGGCTGCTGAGAGCCGTGTCAGAGATGAG 1014
DB      1144 ATTATGAGCTCTGTGTCTGTGATCAGCGCTTGCAGAGGAACCAAGAAAGAGAGCCA 1203
QY      1015 CAGCAGCAGAGAAAGAACTGCTGTGACATTTGAAGATTAAGAACCGGAGAACTTGCAG 1074
DB      1204 CAAAACCGGACAAAAAAGCTCCAGTGCATTTGAAGACAAAGAAAGGAGAACTTTGAG 1263
QY      1075 CGAGCAGTGTGAGCTGAGAAAGCCCGCCAGCGCTTTGAGCAGCAGCGCAAAAGAG 1134
DB      1264 CGAGAAACCTAGATTAAGAAAGAGAGACAGGCACTTTTATGACAGCGCAAAAGAG 1323
QY      1135 CAGAGGCGTGTGCTGCTGAGACCGCGAGAGAGAGAGAAAGCGGAGCGGACAG 1194
DB      1324 CAGAGAGGTGTGCTGCTGAGAAAGGACAGAGCAGAGAGAAAGAAACGAGACGGGAA 1383
QY      1195 GAGCAGAGAGCCAAAGCCGACGCTGAGCTGAGAGAGAGCTGAGAAAGCAGCGGAGCTG 1254
DB      1384 GATCAAGAGCGGAGAGCGGACAGCAAGACTTGGAAAAACAGCTGAGAAACAAACGAACTG 1443
QY      1255 GAGCGCAGCGAGAGAGAGAGAGAGAGAGATCGAGAGCGCGGACCGCAAAACGG 1314
DB      1444 GAGCGGCAACGGGAAAGAGAGCGCAGAAAGAAATCGAAAGAGAGAGCGGCAAAAGA 1503
QY      1315 GAATGGAAGGCGGCAACCTGATGATGGAACGGAACCGGAGACAGGAACCTGCAAT 1374
DB      1504 GAATGAAGAAAGCAACGCGAGTGTGGAAGCAAGAAACAGAAAGCAAGAACTCTTAAT 1563
QY      1375 CAGAGAAACAGAGCAGAGAGGCGACCGTGTCTGAAAGCCAGAGAGAAAGCTCTGAG 1434
DB      1564 CAAGGAATAGAGAGCAAGAGAGCATTTGTCTTAAGCAAGAAAGAAAGAACTCTAGAG 1623
QY      1435 TTGAGTTAAGACTCTGAATGACAAAGACATCAGTACAGAGAAACCTTCAAGATATC 1494
DB      1624 TTGAGCTGGAAGCTCTTAATTAATTAAGAACACAGCTGGAAGAAACCTCCAGATATC 1683
QY      1495 AGTGTGACCTGGCAACCGAGGCAAGAAATGAGAGCAAGAACAGTCTAGAGAGTA 1554
DB      1684 CGGTGTGCTCACTACTCAACGCGACGAATTAAGAAACACTAACAAATCCAGGAACTG 1743
QY      1555 AGAATGCTGAAATCACCACTTACAGCAGCAGATTTGAGAGATCTCAGCAAACTCTTGA 1614
DB      1744 AGAATCCAGAGATCAACACACTTGCACAGCAGAGCTTCAAGATCTCAGCAGCTTGA 1803
QY      1615 AGACTTATTCAGAGAAACAGATATCAATGACCAAGTTAAACAAAGTCCAGCAGAACAT 1674
DB      1804 AAATGATTCCTGAGAAACAGTCCCTTATCGATCAATTTAAAGCAGTTTCAACAAAGCAGT 1863
QY      1675 TTGATAGAGATCGCTTCTTAACTTCAAAAGGCTTGAAGGAAAGAGCTGGCCCG 1734
DB      1864 TTGACAGAGATTCCTTCTTACACTAAAGAGGCAATTTGAACCAAGAGATTTGAGCGC 1923
QY      1735 CAGCAGCTCCGGGAGCAGCTGAGAGAGTGTGAGAGAGACAGAGTCAAAAGCTCAGAG 1794
DB      1924 CAGCAGCTCCGAGATCAAGTTAAGATGAGATGAGAAAGAAAGCGAGCTTACAGAA 1983
QY      1795 ATTGATGTTTCAACAAACAGCTGAGAGAACTGAGAGATATCAATCAACACGAACTC 1854
DB      1984 ATCAGCTAATTAATTAACAAATTAAGAACTGAGAAATTTGTCAACAAAGCAGCACTTC 2043
QY      1855 CAGAAGCAGAGTCTCCGAGAGCGCGCAGCTGAAGGAGAAAGGCGAGAGAGAGAGC 1914
DB      2044 CAGAAACAGAGCTTGTGAACGTGAAGAGTCAAGCAAGAAAGCTGGAACGAAAAACC 2103
QY      1915 CTGAGTTAAGAGAAAGAAAGAAAGAGAGCTCAGAGACGAGTTTCAGAAAGGAGCAAGCAA 1974
DB      2104 AGCAGCTGAGATTAAGCTGAAGAGAGAGAGCAAAAGGAGATGCTGAAGCAGAGAACTT 2163
QY      1975 TGGCTGAGAGATGTGACAGCAGAGAGAGCAGCGCCCGGAAACCCCAAGAGAGAGC 2034
DB      2164 TGGCAGATCGCGTGAAGCAGAGAGAGAGAGAGTAACTTAACTTCAAGATGAGAGAG 2217

```


OY	2035	AACACTGAAGGGGAAAGACAGTCTCGAGAAAGAGAGCGGAAAGAGAGACCAAGCCGGA	2098
Db	2218	AAAGAAAAAAGGAAAGAAATCCGCTCAGAAAGTCGAGGTGAAAAAG-----AAACAAG	2271
OY	2095	ATGCAGAACAAGCAGAGTCGCGCTTTTCCATCCGCATCAAGAGCCAGCTTAAGCTGGCCAC	2154
Db	2272	ATCCAGAGAAACCCACAACCAACCGTTCCACCAACTCCAGAGCCCTGAAAAACTAGAGAGG	2331
OY	2155	CAGGACCCCTGGTCTAACACAGAGAAAGGCCCGCTTACCATTTCGCAACAGAGGTGT	2214
Db	2332	CAGATTCCTGGATGAAATACGAGAAAGCCCTCTTAAGAT---CAACAGGGCCGATGTG	2388
OY	2215	AAAGTGATATATTACCGAGCGCTGATACCCTTTGAAATCCAGAAAGTCAAGATGATCAAC	2274
Db	2389	AAAGTTGTGTATTACCGAGCACTTACCTTTTGAATGCCAGAGTCATGATGAATCACT	2448
OY	2275	ATCCAGCCAGAGATATATAGTCATGGTGTGATGAAGACAGACTGAGAGCCAGATGGCTT	2334
Db	2449	ATTGAGCCCTGGAGCATTTATCATGTGTGATGAAGAACCAACTGAGAGACTGGGTGGCTG	2508
OY	2235	GGAGAGAGCTGAAAAGGGAGACCGGGATGGTTCCTGAAATCTATGCAAAAAAGTTCCA	2394
Db	2509	GGAGGTGAATCTGAAAAGAAAAAGGGCTGGTTTCTGTCCAATGACAGACGAAATGGCC	2568
OY	2395	GAATAATGAGTGTCCCATCTCCAGCCCAACAGTACCGATCTGACATCTGCCCTTCGCCCC	2454
Db	2569	GAGATGTAATTTCCAAAGCAACATTAACAGAGAGCAAACTACAGCCAAACCACTGTCTC	2628
OY	2455	AAACTGGCTCTCGTGAAGACCCCTGCTCTTTGCCAGTGAAGCTCTTCTGAGCCCTCACA	2514
Db	2629	CATGTGGAC-----CTTCCCTGTGCACTGAGGCTTTTCAAAATACATCTTACA	2679
OY	2515	AACCCCAACAACCTGGGAGACTTCAGTTCTCAAGTGCAGCGCCAGAGCTCAAGAGAAAGCCA	2574
Db	2680	AATTCACCAACTGGGACAGACTTCAGCTCAAGGTGGCCAAACCAATTAACCTGATTAAGTA	2739
OY	2575	GAACCGAACAACTGGGATACGTGGGCGGCTCAGCCCTTCTGACCCGTAACCTAGTCTGGC	2634
Db	2740	GAGATGTATTACTGGGATACATGGGCACTCAACCATCTTAACTGTTCCCAAGTGTGT	2799
OY	2635	CAGTTAGGGCAGAGATCAAGCCCTTAACCCACACACAGCACTGGTCTCTCCCACTTCTCC	2684
Db	2800	CAGCACAAGACCGCTCTGTCTTTTACCACAGCACTGTATCTGATCAATGCCCCATCTCT	2859
OY	2695	GTCCTGGCCACGGGTGAAGAGGTGAAAGGGCTACAAAGCGCAAGCCCTGTATCTCCGAGGA	2754
Db	2860	GTACTGTGTCAGGGGTGAAGAAAGTGAAGGCCCTTCAAGCAACAGCCTTGATCTTTGAGGA	2919
OY	2755	GCCAAAAAAGACAACTTAATTTTAACAAAGTGAAGTCATCAACGTTCTGGAAACAG	2814
Db	2920	GCAAAAGAAAGCAACATCTTAATTTTAACAAAATGATGTATTAACCGTCTTGAACAG	2979
OY	2815	CAAGCATGTGGTGTGGAAGAGTTCAAGSTCGAAGGGGTGGTCCCAAGCTTAC	2874
Db	2980	CAGGATATGTGTGTGTTGGGAAAGTTCAAGSTCAAAAGGGCTGTGTTCCCAAAATCTTAT	3039
OY	2875	GTGAATCTCATTTTCAGGGCCCGTAAAGAAATCCAACAGCATCGATCTGGCCCTTCTGA	2934
Db	3040	GTAAGCTTAATATCCGTCCTCACTAAGAAATCTCAACAGCATCGATTCACGTTTCAGAA	3099
OY	2935	AGCTCGTGTAGCTTAAGAGAGTGGCTCCCGGCGCCAGACAGCATTCCTCCGAGGA	2994
Db	3100	AGCCAGGAGTTTAAAGAGAGTCTCTTCAACACAGATTTAAACCTGCATCAAGAGGAA	3159
OY	2995	GAGTTATTTGCATGTACACATACAGAGTTCTGACAAAGAGATTTAACTTTCAACAA	3054
Db	3160	GAATTCATATCATCTATCACATATGAGAGCAACAGGAAAGAGACTTGACCTTTCAAGAA	3219
OY	3055	GGGGAATGATTTGGTGTTCACAAAGAAAGATGCTGTGTGACGGGAACGTTGGCGAC	3114
Db	3220	GGAGATTCGATTTGTGTATTTAAAAAAGATGGCACTGTGTGACCTGGACCGTGTGGAGAG	3279

QY	3115	AAGTCGGAGCTTCCTCTTAACTAATAGTAGGCTTTAAAGATCAAGAGGCTCTGGAACT	3174
Db	3280	AAGATCGGGGTTTTTCCCTCCAACTACGTAGAGGCCAAAGACTCGAGAGCGCGCTGGAAAGC	3339
QY	3175	GCTGGGAAAACAGGAGATTGAGAAAAAACCTGAATTGGCCAGGTTATTGCTTCCAC	3234
Db	3340	GGAGGAAAAGACAGGAAGCTTGGGAAAGAAACAAATAATTGCCAGGTTATTGCACTTAT	3399
QY	3235	GCTGCTACTGGTCCCGAAACACTACCCCTGGCTCTGTGGGACGAGATTCTGATCCGGAAA	3294
Db	3400	GCAGCACTGTCCTTGAACAGCTAACACTGGCTCTTGGCCAACTTATCTTATCCGAAA	3459
QY	3295	AAGAAACCCAGGTGATGGTGGGANGAAGAACTGCAAGCTCGAGGGAAAAAAGCCGCAATA	3354
Db	3460	AAGAACTCAGAGGATGGTGGGAAGAGAGACTCGACAGCAACCTGGGAAAGAAAGCGCAGATT	3519
QY	3355	GGGAGGTTTTCCAGGAAATTATGTCAAACTTCTAGACCCCGGAAACAACAAATACCCCA	3414
Db	3520	GGATGGTTTTCCAGCAAAATTATGTTTAAGCTTTTGAGCCCTGTGACAAATTAAGACACGCCA	3579
QY	3415	ACTGAGCTACCCCAAGACCGCAGTCAGCCAGACAGATGTGCAGATGATCCGAGATACGAT	3474
Db	3580	ACKAAACCCMCCCAACCAACGTCATTCCTCTCTCATGSCCAAGTTATCCGGATGTACGAT	3639
QY	3475	TACACCGCCCAAGACGATGACGAACTGACCTTCAGGCAAGGCCAGATCTATCAAGTCTCTC	3534
Db	3640	TATATCGGCAAGATGACGAGAGATTGACCTTCACTGTAAGGCGAGGTGATTTAACGTGCTG	3699
QY	3535	AACAAAGAGAACCCGGACTGGTGGAAAAGAAAGTCAAGTGGGCAAGTTGGGCTCTTCCCA	3594
Db	3700	AACAAAGAGATCCCGACTGGTGGAAAAGGAAACTGAATGGCCATGTGTGGCTGTCTTCCC	3759
QY	3595	TCCAAATTATGAAAGCTGACCAACAAGATGAGACCCCAAGCCAGCAAT	3640
Db	3760	TCCAACTACGTGAACACTGACCAACAAGATGAGATCTTATGTCAGCAAT	3805

FEATURES	SOURCE
TITLE JOURNAL COMMENT	RESULT 10 BD127168
LOCUS	BD127168
DEFINITION	Primer for synthesizing full-length cDNA and use thereof.
ACCESSION	BD127168
VERSION	BD127168.1 GI:23222113
KEYWORDS	JP 2002017375-A/2599.
SOURCE	Homo sapiens (human).
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 213)
AUTHORS	Ota,T., Nishikawa,T., Isogai,T., Hayashi,K., Ishii,S., Kawai,Y., Makamatsu,A., Sugiyama,T., Naga,K., Kojima,S., Otsuki,T. and Koga,H.
	Primer for synthesizing full-length cDNA and use thereof Patent: JP 2002017375-A 2599 22-JAN-2002; HELIX RESEARCH INSTITUTE
	OS Homo sapiens (human) PN JP 2002017375-A/2599
	PD 22-JAN-2002
	PF 07-JUL-2000 JP 2000253172
	PI TOSHIO OTA,TETSUO NISHIKAWA,TAKAO ISOGAI,KOJI HAYASHI,SHIZUKO PI ISHII, PI YURI KAWAI,AI WAKAMATSU,TOMOYASU SUGIYAMA,KEIICHI NAGAI, PI SHINCHI KOJIMA,
	PI TETSUJI OTSUKI,HISASHI KOGA
	PC
	C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/ PC 10, PC C12P21/02,C12Q1/68//C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC Primer for synthesizing full-length cDNA and use thereof FH Key Location/Qualifiers (205)..(2130). FT CDS Location/Qualifiers 1..2131

```

/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT      694 a      469 c      526 g      442 t
ORIGIN
Query Match      40.7%; Score 1483.8; DB 6; Length 2131;
Best Local Similarity 86.0%; Pred. No. 0;
Matches 1657; Conservative 0; Mismatches 267; Indels 3; Gaps 1;

Oy 1 ATGGCTCAGTTTCCACACCTTTGCGTGTGATGCTGTGAGGATGATGAGTGTGAG 60
Db 205 ATGGCTCAGTTTCCACACCTTTGCGTGTGAGGATGATGAGTGTGAGTGTGAG 264
Oy 61 GAAAGGCGCAAGCATGACGACGAGTTCCTTAGCCGGAAGCCGATAGCGGATTTTACT 120
Db 265 GAAAGGCGCAAGCATGACGACGAGTTCCTTAGCCGGAAGCCGATAGCGGATTTTACT 324
Oy 121 GGTGATCAAGGAGGAACTTTTTCATCTGGGTTACCTCAGCCTGTCTTAGCAAA 180
Db 325 GGTGATCAAGGAGGAACTTTTTCATCTGGGTTACCTCAGCCTGTCTTAGCAAA 384
Oy 181 ATATGGGCGTACGCGACATGATACATGAGATGAGATGATCAAGTGAATTTTCATA 240
Db 385 ATATGGGCGTACGCGACATGATACATGAGATGAGATGATCAAGTGAATTTTCATA 444
Oy 241 GCGATGAAGCTTATCAAACTGAAAGTACAAAGATTCAGCTCCCTCCACATTTCCCT 300
Db 445 GCGATGAAGCTTATCAAACTGAAAGTACAAAGATTCAGCTCCCTCCACATTTCCCT 504
Oy 301 GTCATGAACAGCAACAGTGGCTATTTCCAGTGCACAGCACTTTGTATGAGAGGAT 360
Db 505 GTCATGAACAGCAACAGTGGCTATTTCCAGTGCACAGCACTTTGTATGAGAGGAT 564
Oy 361 GCTAGCATGCCACCACTCAGCTGTGCTCTGTCGCAATGGGCTCCATTCAGATTGT 420
Db 565 GCTAGCATGCCACCACTCAGCTGTGCTCTGTCGCAATGGGCTCCATTCAGATTGT 624
Oy 421 GGAATGTCCTCAACCTTATCTCTGCTCCCTCAGACAGAGTGTCTCCCTGCTTAC 480
Db 625 GGAATGTCCTCAACCTTATCTCTGCTCCCTCAGACAGAGTGTCTCCCTGCTTAC 684
Oy 481 GGGGCTCTCCCGTATCAGGCTCTGCTGCTGGTGGCATCTGACAGCCATGGGCA 540
Db 685 GGGGCTCTCCCGTATCAGGCTCTGCTGCTGGTGGCATCTGACAGCCATGGGCA 744
Oy 541 AAGAGTTCCTCTCAGACATCTGTGTCAGGGGTCACAATTAACACTAATTACAGAG 600
Db 745 AAGAGTTCCTCTCAGACATCTGTGTCAGGGGTCACAATTAACACTAATTACAGAG 804
Oy 601 GCACATCATTTGATGTGCGCCAGCGCCCTCCAGACAGAGATGGGCTGTGCTCAGTCA 660
Db 805 GCACATCATTTGATGTGCGCCAGCGCCCTCCAGACAGAGATGGGCTGTGCTCAGTCA 864
Oy 661 TCAAGGCTGAATTAAGGAGTTTCAACAGCCGCAAAACTATAGTGGAGCACTTA 720
Db 865 TCAAGGCTGAATTAAGGAGTTTCAACAGCCGCAAAACTATAGTGGAGCACTTA 924
Oy 721 ACAGGTCCTCCAGGCAAGAACTATCTCATGCAATCAAGTTTACCCAGGCTCAGCTGCT 780
Db 925 ACAGGTCCTCCAGGCAAGAACTATCTCATGCAATCAAGTTTACCCAGGCTCAGCTGCT 984
Oy 781 TCAATATGGAATCTTTCTGACATTTGATCAAGTGAAGAACTCACTGCGAAGAAATTTATC 840
Db 985 TCAATATGGAATCTTTCTGACATTTGATCAAGTGAAGAACTCACTGCGAAGAAATTTATC 1044
Oy 841 CTAGCTATGCACTAATGATGTGCANGTGTGCGACGCACTGCGCGGCTCTGCTGCT 900
Db 1045 CTAGCTATGCACTAATGATGTGCANGTGTGCGACGCACTGCGCGGCTCTGCTGCT 1104
Oy 901 CCAGATATCATCTCTCTCTCTCAGAAAGATTGCTCCGCGAGTGGAGTGTCCGTCAATA 960
Db 1105 CCAGATATCATCTCTCTCTCTCAGAAAGATTGCTCCGCGAGTGGAGTGTCTGTCAATA 1164

```

```

Oy 961 AGCTCTTCTTCTGTGATGATGAGGCTGCTGAGGAGCGTGTGACAGAGATGACAGCAG 1020
Db 1165 AGCTCAACATCTGTAGGTGACAGGCTTACAGAGGAAACAGTTTATGAAATGAAACAA 1224
Oy 1021 C---CAGAGAAAGAACTGCTGTGATCTTGAAGATTAAGAGCGGAGAACTTTCAGCGA 1077
Db 1225 CAATTAGAAAGAAATTAACCTGTAACTTTGAAGATTAAGAAAGCGGAGAACTTTGAACGT 1284
Oy 1078 GGCAGTGTGAGCTGTGAGAAAGCCGCCAAGCGCTCTTGGAGCAGCAGCCGAAAGAGCAG 1137
Db 1285 GGCAGCTGTGAACTGTGAGAAAGCAAGGCAAGCTCTCTGTGAAACAGCAGCCGAAAGAGCAG 1344
Oy 1138 GAGGCGTTGGCTGAGCTGTGAGAGCGCGGAGAGAGAGAAAGCGGAGAGCGCAGAG 1197
Db 1345 GAGGCGCTGGCCAGCTGTGAGAGCGGCGGAGAGAGAGAGAGAGAGAGCTGAGCAGAG 1404
Oy 1198 CAGAGAGCCAAAGCCGAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1257
Db 1405 CAAGAGGCGCAAAAGAAAGCACTGTGAAGAGCACTGTGAAGAGCAAGCGGAGCTAGAA 1464
Oy 1258 CCGCAGAGAGAGAGAGAGAGAGAGAGAGAGATCGAGAGGCGGAGCGCAAAACGGGAA 1317
Db 1465 CCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1524
Oy 1218 CTGGAAGGCGAGCAACTTGAATGGGAACGGAACCGGAGAGAGAGAGAGAGAGAGAGAG 1377
Db 1525 CTGGAAGGCGAGCAACTTGAATGGGAACGGAATGGAGAGAGAGAGAGAGAGAGAGAGAG 1584
Oy 1378 AGGAACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1437
Db 1585 AGGAACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1644
Oy 1438 GAGTTGAAGCTGTGAATGACAAAGAGATGAGTGAAGAGAGAGAGAGAGAGAGAGAGAG 1497
Db 1645 GAGTTGAAGCTGTGAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1704
Oy 1498 TGTGAGCTGCGAACCAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1557
Db 1705 TGTGAGCTGCGAACCAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1764
Oy 1558 ATTGCTGAATCAACCACTTACAGCAGCAGAGTTCAGAGATCTCAGCAATGCTTGAAGA 1617
Db 1765 ATTGCTGAATCAACCACTTACAGCAGCAGAGTTCAGAGATCTCAGCAATGCTTGAAGA 1824
Oy 1618 CTTATTTCCAGAGAAACAGATCTCAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1677
Db 1825 CTTATTTCCAGAGAAACAGATCTCAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1884
Oy 1678 CATAGAGAGCTGCTCTTACCTCAAAAGAGCTTGGAGAGAGAGAGAGAGAGAGAGAGAGAG 1737
Db 1885 CATAGAGAGCTGCTCTTACCTCAAAAGAGCTTGGAGAGAGAGAGAGAGAGAGAGAGAGAG 1944
Oy 1738 CAGCTCCGGAGAGCAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1797
Db 1945 CAGCTCCGGAGAGCAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2004
Oy 1798 GATGTTTCAACAACAGCTGAGAGAGAGAGAGAGAGATACATAGCAACAGCAACTCCAG 1857
Db 2005 GATGTTTCAACAACAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2064
Oy 1858 AAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1917
Db 2065 AAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2124
Oy 1918 GAGTTAG 1924
Db 2125 GAATTAG 2131

```

```

RESULT 11
AK074554
LOCUS      AK074554      2131 bp      mRNA      linear      PRI 03-SEP-2002

```



```

DEFINITION Homo sapiens cDNA FLJ90073 fis, clone HEMBA1004110, highly similar
              to Homo sapiens interectin short form mRNA.
ACCESSION   AK074554
VERSION     AK074554.1 GI:22760070
KEYWORDS    oligo capping, fis (full insert sequence).
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
REFERENCE   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
AUTHORS     Isogai,T., Ota,T., Nishikawa,T., Hayaishi,K., Otsuki,T.,
            Sugiyama,T., Suzuki,Y., Nagai,K., Sugano,S., Ishii,S.,
            Kawai-Hio,Y., Saito,K., Yamamoto,J., Wakamatsu,A., Nakamura,Y.,
            Kojima,S., Nagahara,K., Masuko,Y., Ono,T., Okano,K., Yoshikawa,Y.,
            Aotsuma,S., Sasaki,N., Hattori,A., Okumura,K., Iwayanagi,T. and
            Nimomiya,K.
TITLE       NEDO human cDNA sequencing project
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 2131)
AUTHORS     Isogai,T. and Otsuki,T.
TITLE       Direct Submission
JOURNAL     Submitted (25-MAR-2002) Takao Isogai, Helix Research Institute,
            Genome Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            (E-mail:genome@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
            NEDO human cDNA sequencing project supported by Ministry of
            Economy, Trade and Industry of Japan; cDNA full insert sequencing:
            Research Association for Biotechnology; cDNA library construction:
            Institute of Medical Science, University of Tokyo, Laboratory of
            Genome Structure, Human Genome Center; cDNA 5'- & 3'-end one pass
            sequencing and clone selection: Helix Research Institute (supported
            by Japan Key Technology Center etc.).
FEATURES
    source
        1..2131
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="HEMBA1004110"
            /cvsave_type="whole embryo, mainly head"
            /clone_lib="HEMBA1"
            /dev_stage="embryo, 10 weeks"
            /note="Cloning vector: pMB18SFL3"
BASE COUNT  694 a 469 c 526 g 442 t
ORIGIN
Query Match      40.7%; Score 1483.8; DB 9; Length 2131;
Best Local Similarity 86.0%; Freq. No. 0;
Matches 1657; Conservative 0; Mismatches 267; Indels 3; Gaps 1;
QY 1 ATGCTCAGTTTCCACACCTTTCGCTGAGCTGAGTCTGGCCATTAACCTGTGAG 60
DB 205 ATGGCTCAATTTCCACACCTTTCGCTGAGCTGAGTCTGGCCATTAACCTGTGAG 264
QY 61 GAAAGGGCAAGCATGACGAGCATGTTCTTACCTGAGCCGATACGGGATTTATTA 120
DB 265 GAAAGGGCAAGCATGATGACGAGCATGTTCTTAAAGCAATATCTGGATTCATTA 324
QY 121 GGGATCAAGCCAGGAATTTTTTTTCCATCTGGGTACTAGCTGTCTTAAAGCA 180
DB 265 GAAAGGGCAAGCATGATGACGAGCATGTTCTTAAAGCAATATCTGGATTCATTA 324
QY 325 GGGATCAAGCCAGGAATTTTTTTTCCATCTGGGTACTAGCTGTCTTAAAGCA 384
DB 181 ATATGGGCGCTAGCGGACATGATTAAGATGAAGATGATCAAGTGAATTTTCCATA 240
QY 385 ATATGGGCGCTAGCGGACATGATTAAGATGAAGATGATCAAGTGAATTTTCCATA 444
DB 241 GCCATGAAGCTTATCAAACTGAAGTACAAAGATATCACTCCCTCCACACTTCCCT 300
QY 445 GCTATGAAGCTTATCAAACTGAAGTACAAAGATATCACTCCCTCCACACTTCCCT 504
DB 301 GTCATGAAGCAGAACCAAGGCTATTTTCCATGTCAGCAGCATTTTGTATAGAGGATT 360
QY 505 GTCATGAAGCAGAACCAAGGCTATTTTCCATGTCAGCAGCATTTTGTATAGAGGATT 564
DB 361 GTCATGAAGCAGAACCAAGGCTATTTTCCATGTCAGCAGCATTTTGTATAGAGGATT 420

```

```

DB 565 GCAGAGATGCCAACCGCTTACAGCTGTGTCTCCAGTGCCAAATGGATTCATTCCAGTTGTT 624
QY 421 GGAATGTCTCCACCTTATGATCTTGTGCTCCAGACAGAGTGCTCCCTGGCTAAC 480
DB 625 GGAATGTCTCCACCTTATGATCTTGTGCTCCAGACAGAGTGCTCCCTGGCTAAC 684
QY 481 GGGGCTCCCGGCTATACAGCTGCTGCTGGCTTGGCAATCTGCACCCCAATGGCCA 540
DB 685 GGGGCTCCCGGCTATACAGCTGCTGCTGGCTTGGCAATCTGCACCCCAATGGCCA 744
QY 541 AAGAGTCTTCTCTTCCAGCAGATCTGCTGAGGGTCACAATTAACACTAAGTTACAGA 600
DB 745 AAGAGTCTTCTCTTATGATGATCTGCTGAGGGTCACAATTAACACTAAGTTACAGA 804
QY 601 GCACATATCTTGATGTGCCAGCGCCCTCCAGACAGAGAAATGGGCTGTGCTCAGTCA 660
DB 805 GCACATATCTTGATGTGCCAGCGCTCCAGACAGAGAAATGGGCTGTGCTCAGTCA 864
QY 661 TCAGGCTGAATATACAGGAGTATTTCAACAGCCACGCAAAACATATGATGACACTTA 720
DB 865 TCAGGCTGAATATACAGGAGTATTTCAATATGATGATGACAAACATATGATGACACTTA 924
QY 721 ACAGGTCCCGCAGCAAGAACTATTTCTCATGCAATCAAGTTTACCCAGGCTCAGCTGCT 780
DB 925 ACAGGTCCCGCAGCAAGAACTATTTCTCATGCAATCAAGTTTACCCAGGCTCAGCTGCT 984
QY 781 TCATATGGAATCTTCTTGCATTTGATATCAATCAATGAGAAATCTCATGCAAGAAATTTATC 840
DB 985 TCATATGGAATCTTCTTGCATTTGATATCAATCAATGAGAAATCTCATGCAAGAAATTTATC 1044
QY 841 CTAGCTAGCACTTAATGATGATTTGATGATGCTGCTGAGCCCTGCGCGCTCTGCTCT 900
DB 1045 CTAGCTAGCACTTAATGATGATTTGATGATGCTGCTGAGCCCTGCGCGCTCTGCTCT 1104
QY 901 CCAGAAATACATCCCTCTCTCTCTTCTTCAAGAGTGTGCTCCGAGTGGAGTGTCCGTCATA 960
DB 1105 CCAGAAATACATCCCTCTCTCTCTTCTTCAAGAGTGTGCTCCGAGTGGAGTGTCCGTCATA 1164
QY 961 AGCTCTTCTCTTGTGATGATGAGAGGCTGCTGAGAGCGCTGCTCAAGAGATGAGACGAG 1020
DB 1165 AGCTCTTCTCTTGTGATGATGAGAGGCTGCTGAGAGCGCTGCTCAAGAGATGAGACGAG 1224
QY 1021 C---CAGAGAGAACTGCTGCTGATGATTTGAAGTAAAGACGGGAGAACTTCCAGCGA 1077
DB 1225 CAAATTAAGAAAGAAATTAACCTGTAACTGTTGAAGTAAAGACGGGAGAACTTCCAGCGA 1284
QY 1078 GGCAGTGTGAGCTGAGAGAGGCGCCCAAGGCTCTTGGAGCAGACAGGCAAGAGCGAG 1137
DB 1285 GGCAGTGTGAGCTGAGAGAGGCGCCCAAGGCTCTTGGAGCAGACAGGCAAGAGCGAG 1344
QY 1138 GAGCGGTTGGCTCAAGCTGAGAGCGCGCAGCAGAGAGAGAAAGACGGGAGCGCCAGAG 1197
DB 1345 GAGCGGTTGGCTCAAGCTGAGAGCGCGCAGCAGAGAGAGAAAGACGGGAGCGCCAGAG 1404
QY 1199 CAGGAGGCCAAGCGGAGCTGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1257
DB 1405 CAGGAGGCCAAGCGGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1464
QY 1258 CGGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1317
DB 1465 CGGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1524
QY 1318 CTGAAAGGCAAGCAACTTGAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1377
DB 1525 CTGAAAGGCAAGCAACTTGAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1584
QY 1378 AGGAACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1437
DB 1585 AGGAACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1644
QY 1438 GAGTAGAAGAGCTGTAATGACAAAGAGATCAGTAGAAGAGAGAGAGAGAGAGAGAGAG 1497

```

```

Db      1645 GAATTAGAGCTCTAATATGATTAAGCATCACTAGAGGAAAATTCAAGATATCAGA 1704
Qy      1498 TGTGACCTGCGACCCAGAGCAAGAAATTTGAGACGAGCAACTCTAGAGGTAGA 1557
Db      1705 TGTGATTTGACACCCAAAGGAGAAATTTGAGACGACAAACAATCTAGAGGTAGA 1764
Qy      1558 ATTGCTGAATACCCACTTACAGCAGAGTTCAGAAATCTCAGCAATCTTGGAGA 1617
Db      1765 ATTCGCAATACCCACTTACAGCAGCAATTTACGAAATCTCAGCAATCTTGGAGA 1824
Qy      1618 CTATTCAGAGAAACAGATACTCAGTACGACGATTAAACAGTCCAGCAGAACTTTG 1677
Db      1825 CTATTCAGAGAAACAGATACTCAGTACGACGATTAAACAGTCCAGCAGAACTTTG 1884
Qy      1678 CATGAGACTGCTCTTCACTTACCTCAAGAGCTTGGAGCAAGAGACTGGCCGCGAG 1737
Db      1885 CACAGAGTTCACTTGTACACTTAAAGAGCTTTAGAGCAAAAGAACTGCTGCGAG 1944
Qy      1738 CAGCTCCGAGCAGACTGAGAGAGTGGAGAGAGACGAGTCAAGCTGAGAGATT 1797
Db      1945 CACCTAGAGACCAACTGAGTGAAGTGAAGAAAGAACTAGATCAAACTACAGAGATT 2004
Qy      1798 GATGTTTCAACACACGACTGAGAGACTGAGAGATACATAGCAACAGCACTCCAG 1857
Db      2005 GATATTTCAATATACAGCTGAGAGAACTAAGAGAAATACAAATACCAACTCCAG 2064
Qy      1858 AAGCAGAGTCCCTGAGAGCGGAGTGAAGAGCAAGAGCAGAGAGAGAGAGCTG 1917
Db      2065 AAGCAAAAGCTCATGAGAGCTGAAGACTGAACGAAAGAAAGCAAGAAAGATCAT 2124
Qy      1918 GAGTTAG 1924
Db      2125 GAATTAG 2131

RESULT 12
BD158570      2131 bp      DNA      linear      PAT 17-JAN-2003
LOCUS      BD158570
DEFINITION      Primer for synthesizing full-length cDNA and use thereof.
ACCESSION      BD158570.1 GI:27864328
VERSION      JP 2002191363-A/13413.
KEYWORDS      Homo sapiens (human)
SOURCE      Homo sapiens
ORGANISM      Homo sapiens
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
TITLES      1 (bases 1 to 2131)
JOURNAL      Ota,T., Isegai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,
      Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.
      Primer for synthesizing full-length cDNA and use thereof
      Patent: JP 2002191363-A 13413 09-JUL-2002;
      HELIX RESEARCH INSTITUTE
COMMENT      OS Homo sapiens (human)
      PN JP 2002191363-A/13413
      PD 09-JUL-2002
      PF 28-JUL-2000 JP 2000280990
      PI TOSHIO OTA,TAKAO ISEGAI,TETSUO NISHIKAWA,KOJI HAYASHI,KOJI
      PI SAITO,KOJI YAMAMOTO,SHIZUKO ISHII,TOMOYASU SUGIYAMA,AI WAKAMATSU,
      PI KEIICHI NAGAI,TETSUJI OTSUKI
      PC
      C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/
      10,C12P21/02,C12Q1/68//C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC
      Primer for synthesizing full-length cDNA and use thereof FH Key
      Location/Qualifiers
      FT CDS      Location/Qualifiers
      source      1..2131
      /organism="Homo sapiens"
      /mol_type="genomic DNA"
      /db_xref="taxon:9606"
BASE COUNT      691 a      448 c      572 g      420 t

```

```

ORIGIN
Query Match      34.6%; Score 1259.2; DB 6; Length 2131;
Best Local Similarity 81.2%; Pred. No. 9.1e-275;
Matches 1539; Conservative 0; Mismatches 243; Indels 114; Gaps 2;

Qy      1 ATGGCTAGTTTCCACACCTTTGCGGTAGCTGTGATCTGCGCATTAACCTGTGAG 60
Db      347 ATGCTCAGTTTCCAAACCTTTTGGTGGCAGCTGGATATCTGGGCATTAACCTGTGAG 406
Qy      61 GAAAGGCCAAGCATGACAGCAGCTTCTTAGCCTGAAGCCGATAGCGGATTTATTACT 120
Db      407 GAAAGGCCAAGCATGATGATCAGCTTCAATAGTTTAAAGCAATATCTGATTCATTACT 466
Qy      121 GGTATCAACCGGAACTTTTTCATCTGGGTACTCTGACCTGCTTAGCAAC 180
Db      467 GGTATCAACCTGAAACCTTTTTCATCTGGGTACTCTGACCTGCTTAGCAAC 526
Qy      181 ATATGGCGCTAGCGGACATGAATTAACGATGAAGATGATCAAGTGAATTTTCATA 240
Db      527 ATATGGCGACTAGCTGACATGAATTAATGATGAAGATGATCAAGTGAATTTTCATA 586
Qy      241 GCATGAGCTTATCAACTGAGCTACAGCTACAAAGATATAGCTCCCTCCACACTTCCCT 300
Db      587 GCTATGAACCTTATCAACTGAGCTACAAAGATATAGCTACCTCTGACACTTCCCT 646
Qy      301 GTCATGAACAGCAACAGTGGCTATTTCAGTGCACAGACTTGGTATAGAGGAGATT 360
Db      647 GTCATGAACAGCAACAGTGGCTATTTCAGTGCACAGACTTGGTATAGAGGAGATT 691
Qy      361 GCTAGCATGCGCACCACTCAGAGCTGTGCTCTGTGCAATGGGCTCCATTCCAGTTGTT 420
Db      692 ----- 691
Qy      421 GAAATGTCACACCTTAGTATCTTGTGCTCCACAGCAGAGTCTCCCTGGCTAAC 480
Db      692 -----SCAGTGTGCCCCCTGCTGCTAAC 715
Qy      481 GGGGCTCTCCGCTCATACAGCCCTGCTGCTTGGCGATCTGAGCAGACATGGCCA 540
Db      716 GGGGCTCTCCGCTGTATACACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 775
Qy      541 AAGAGTCTTCTTCCAGCAGATCTGCTCAGAGGTCACAAATTAACATAATTACAGAG 600
Db      776 AAGAGTCTTCTTCTTATGATATCTGCTCAGAGGTCACAAATTAACATAATTACAGAG 835
Qy      601 GCACATATCATTCATGTCGCCAGCGCCCTCCAGCAGCAGAAATGGCTGCTCAGTCA 660
Db      836 GCACATATCATTCATGTCGCCAGCGCTCCACAGTGGCAGAGTGGCTGCTCAGTCA 895
Qy      661 TCAAGCTGAAATTCAGAGGATTATCAACAGCCAGCAAAATATAGAGTGAAGCACTTA 720
Db      896 TCAAGCTGAAATTCAGAGGATTATCAACAGCTGATATCAAAATATAGAGTGAAGCACTTA 955
Qy      721 AAGAGTCCCGCAGCAAAATCTTCTCATGCAATCAAGTTTACCCAGGCTCAGCTGACT 780
Db      956 AAGAGTCCCGCAGCAAAATCTTCTCATGCAATCAAGTTTACCCAGGCTCAGCTGACT 1015
Qy      781 TCAATATGAAATCTTCTGATTCATGATCAAGATGAAAACCTCAGCAGAGAAATTTATC 840
Db      1016 TCAATATGAAATCTTCTGATTCATGATCAAGATGAAAACCTTACAGCAGAGAAATTTATC 1075
Qy      841 CTACCTATGACCTTAATTTGATGTTGTCATGCTGCTGTCAGCCGCGCTGCTGCT 900
Db      1076 CTGCAATGACCTCATTGATGATGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCT 1135
Qy      901 CCAATATACATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 960
Db      1136 CCAATATACATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1195
Qy      961 AGCTCTTCTCTCTGATTCAGAGGCTGCTGAGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020
Db      1196 AGCTCAACATCTGATGATCAGAGGCTACAGAGGAAACAGTTTATAGAGATGAACAACAA 1255

```

QY 1021 C---CAGAGAGAACTGCTGTGACATTTGAAGATPAGAAGCGGAGAACTTCGAGCGA 1077
 Db 1256 CAATTGAAAAGAAATTACTGTAAACGTTTGAAGATPAGAAGCGGAGAACTTTGAACGT 1315
 QY 1078 GGGAGTGTGAGCTGAGAAAGCCCGCCAAAGCCCTCTTGAGACAGACGCGAAAGAGCAG 1137
 Db 1316 GGGACCTGTGAACCTGAGAAACGAAAGCAAGCTCTCTGGAACAGCAGCGCAAGAGCAG 1375
 QY 1138 GAGCGGTGGCTGAGCTGAGCGCGCGAGACAGAGAGAAAGAGCGGAGCGCCGAGAG 1197
 Db 1376 GAGCGCTGGCCAGCTGAGCGCGCGAGACAGAGAGAGAGAGCGGTGAGCGCCGAGAG 1435
 QY 1198 CAGAGAGCCAAAGCGGAGCTGAGCTGAGAAACAGCTGAGAGAGCAGCGGAGCTGAG 1257
 Db 1436 CAGAGCGCGAAAGAAACAAGCTGAGAAACAGCTGAGAAACAGCGGAGCTGAGAA 1495
 QY 1258 CGGACCGAGAGAGAGAGAGAGAGAGAGATGAGAGCGCGAGCGCGAAACCGGAA 1317
 Db 1496 CAGCAGAGAGAGAGAGAGAGAGAGAAATGAGAGCGAGAGCGCTGCGAAACCGGAA 1555
 QY 1318 CTGGAAGAGCAGAGCAACTTGAATGGGAACCGGAGACGAACTCTGAGATCAG 1377
 Db 1556 CTGGAAGAGCAGAGCAACTTGAATGGGAACCGGAGACGAACTCTGAGATCAG 1615
 QY 1378 AGGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1437
 Db 1616 AGGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1675
 QY 1438 GAGTTAGAAAGCTGTGAATGCAAAAAGCATGAGCTGAGAGAGAGAGAGAGAGAG 1497
 Db 1676 GAGTTAGAAAGCTGTGAATGCAAAAAGCATGAGAGAGAGAGAGAGAGAGAGAGAG 1735
 QY 1498 TGTGAGCTGGCAACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1557
 Db 1736 TGTGAGCTGGCAACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1795
 QY 1558 ATTGCTGAATTCACCACTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1617
 Db 1796 ATTGCTGAATTCACCACTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1855
 QY 1618 CTTATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1677
 Db 1856 CTTATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1915
 QY 1678 CATTAGAGACTCGCTTCTTACCTTCAAAAAGAGCTTGAAGAGAGAGAGAGAGAG 1737
 Db 1916 CATTAGAGACTCGCTTCTTACCTTCAAAAAGAGAGAGAGAGAGAGAGAGAGAGAG 1975
 QY 1738 CAGCTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1797
 Db 1976 CAGCTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2035
 QY 1798 GATGTTTCAACAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1857
 Db 2036 GATGTTTCAACAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2095
 QY 1858 AAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1893
 Db 2096 AAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2131
 RESULT 13
 AK027846 2131 bp mRNA linear PRI 01-AUG-2002
 LOCUS AK027846
 DEFINITION Homo sapiens cDNA FLJ14940 fis, clone PLACE1010942, highly similar
 to Homo sapiens intersecin long isoform (ITSN) mRNA.
 ACCESSION AK027846.1 GI:14042823
 VERSION AK027846.1
 KEYWORDS oligo capping; fis (full insert sequence).
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE
 1
 AUTHORS Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
 Nishikawa,T., Nagai,K., Sugano,S., Ishibashi,T., Fujimori,K.,
 Tanai,H., Kimata,M., Watanabe,M., Hirooka,S., Ishii,S., Kawai,Y.,
 Saito,K., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagahori,K.,
 Masuko,Y. and Kanehori,K.
 NEDO human cDNA sequencing project
 TITLE
 JOURNAL
 2 (bases 1 to 2131)
 REFERENCE Isogai,T. and Otsuki,T.
 AUTHORS Isogai,T. and Otsuki,T.
 DIRECT SUBMISSION
 JOURNAL
 Submitted (10-MAY-2001) Takao Isogai, Helix Research Institute,
 Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
 (E-mail:genomics@ri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
 NEDO human cDNA sequencing project supported by Ministry of
 Economy, Trade and Industry of Japan; cDNA full insert sequencing;
 Research Association for Biotechnology; cDNA library construction;
 5'- & 3'-end one pass sequencing and clone selection; Helix
 Research Institute (supported by Japan Key Technology Center etc.)
 and Department of Virology, Institute of Medical Science,
 University of Tokyo.
 FEATURES
 Location/Qualifiers
 1..2131
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="PLACE1010942"
 /tissue_type="Placenta"
 /clone_1b="PLACE1"
 /note="Cloning vector: pME18FL3"
 BASE COUNT 691 a 448 c 572 g 420 t
 ORIGIN
 Query Match 34.6%; Score 1259.2; DB 9; Length 2131;
 Best Local Similarity 81.2%; Pred. No. 9.1e-275;
 Matches 1539; Conservative 0; Mismatches 243; Indels 114; Gaps 2;
 QY 1 ATGGCTCAAGTTTCCACACCTTTGGGTGATGCTGAGATGTTGGGCAATACGTGGAG 60
 Db 347 ATGGCTCAAGTTTCCACACCTTTGGGTGATGCTGAGATGTTGGGCAATACGTGGAG 406
 QY 61 GAAAGGGCAAGCATGAGACGAGAGTTCCTTACCTGGAACGAGATGCGGATTTATTA 120
 Db 407 GAAAGGGCAAGCATGAGATGAGAGTTCCTTACCTGGAACGAGATGCGGATTTATTA 466
 QY 121 GGTGATCAAGCAGAGAACTTTTTCATCTGAGTTTACTGAGCTGCTTACGACAA 180
 Db 467 GGTGATCAAGCAGAGAACTTTTTCATCTGAGTTTACTGAGCTGCTTACGACAA 526
 QY 181 ATATGGGCGCTTACGAGCATGATTAACATGAGAGATGATCAAGTGAATTTTCATA 240
 Db 527 ATATGGGCGCTTACGAGCATGATTAACATGAGAGATGATCAAGTGAATTTTCATA 586
 QY 241 GCCATGAAGCTTATCAAACTGAAGTACAGGATATACGCTCCCTCCACACTCCCT 300
 Db 587 GCTATGAACCTTATCAAACTGAAGTACAGGATATACGCTCCCTCCACACTCCCT 646
 QY 301 GTCATGAACACAGCAACAGTGTGCTATTTTCAGTGAACAGCAATTTGATAGAGGAT 360
 Db 647 GTCATGAACACAGCAACAGTGTGCTATTTTCAGTGAACAGCAATTTGATAGAGGAT 691
 QY 361 GCTAGAGTCCACCACTACAGCTGTGCTCTGTGCAATGGGCTTCATTCAGTTGTT 420
 Db 692 GCTAGAGTCCACCACTACAGCTGTGCTCTGTGCAATGGGCTTCATTCAGTTGTT 691
 QY 421 GGAATGTCTCACACCTTAGTATCTCTGCTCCAGAGAGAGAGTCCCTCCGCTTAC 480
 Db 692 GGAATGTCTCACACCTTAGTATCTCTGCTCCAGAGAGAGAGTCCCTCCGCTTAC 715
 QY 481 GGGGCTCTCCGCTATACAGGCTGTGCTGCTTGGGATCTGCAAGCCATGAGGCA 540
 Db 716 GGGGCTCTCCGCTATACAGGCTGTGCTGCTTGGGATCTGCAAGCCATGAGGCA 775

```

QY 541 AAGAGTTCTTCTCAGAGAGATCTGTCAGGGTCAACAATTAAACCTAAGTTACAGAG 600
DB 776 AAGAGTTCTTCTTGTAGAGATCTGTCAGGGTCAACAATTAAACCTAAGTTACAGAG 835
QY 601 GCACAAATCATTCATGTCGCGAGCGCCCTCCAGAGAGAAATGGCTGTGCTCAGTCA 660
DB 836 GCACAGATTTGATGTCGCGAGTGTCCACAGTGGGAGAGTGGGCTTCTCAGTCA 895
QY 661 TCAGAGCTGAAATACAGGAGTGTATTCAACAGCCAGCAAAACTATATAGTGAACATTA 720
DB 896 TCAGAGCTGAAATACAGGCAATTAATCATATGTCAGCAAAACTATATAGTGAACATTA 955
QY 721 ACAGGTCGCCAGGAGAGAACTATTCTCATGCAATCAATTTAACCCAGGCTCAGTGGCT 780
DB 956 ACAGGTCGCCAGGAGAGAACTATTCTCATGCAATCAATTTAACCCAGGCTCAGTGGCT 1015
QY 781 TCAATATGGAATCTTTCTGACATTCATGATCAAGATGAGAAACTCATCTGAGAAATTTATC 840
DB 1016 TCAATATGGAATCTTTCTGACATTCATGATCAAGATGAGAAACTCATCTGAGAAATTTATC 1075
QY 841 CTAGTATGACCTTAATGATGTCCTCATGTCGTGTCAGCCACTGCCCGCTCTGCT 900
DB 1076 CTGAGCAATGCACTCATTTGATGATGTCATGTCGCGCAACCACTGCTCTGCT 1135
QY 901 CCAGAAATCATCCCTCTTCTTCTGAGAGAGTTCGCTCGGAGTGGAGTGTCCGTCATA 960
DB 1136 CCAGAAATCATTCACCTTCTTCTTGAAGAGTTCATCTGGAGTGGATGTCATCTGTCATA 1195
QY 961 AGCTCTTCTTCTGTCATCAGAGGCTGCTGAGAGCCGTCGTCAGAGATGAGCAGCAG 1020
DB 1196 AGCTCAATCATCTGTATGATCAGAGGCTTACAGAGAACAGTTTATGAAGATGAACAACA 1255
QY 1021 C---CAGAGAGAACTGCTGTGTCATTTGAAGTAAAGACGGGAGAACTTTCAGAGCA 1077
DB 1256 CAATTAGAGAAAGAAATTACTGTATACGTTTGAAGTAAAGACGGGAGAACTTTCAGAGCA 1315
QY 1078 GGCAGTGTGAGCTGAGAGAGCGCGCAAGCGCTCTTGGAGCAGCAGCGCAAGAGAGAG 1137
DB 1316 GGCAGCTGGAATCTGAGAGAGAGAGAGAGCTCTTCTGAGAGCAGCAGCGCAGAGAGCAG 1375
QY 1138 GAGCGGTTGCTCAGCTGAGAGCGCGCCAGAGCAGAGAGAAAGACGGGAGCGCCAGAGAG 1197
DB 1376 GAGCGGCTGCGCCACTGAGAGCGCGCGAGAGCAGAGAGAGAGAGAGAGAGCGCCAGAGAG 1435
QY 1198 CAGAGGCGCAGAGCGGAGCTGAGAGTGAAGAGCAGCTGAGAGAGAGCGGAGAGCTGAGAG 1257
DB 1436 CAGAGGCGCAGAAAGCAACTGAGAGCTGAGAGAGCAACTGAGAAAGCAGCGGAGAGCTAGAA 1495
QY 1258 CGGAGCGAGAGAGAGAGAGAGAGAGATGAGAGAGCGCGAGCGCGCAAAAGCGGAG 1317
DB 1496 CAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1555
QY 1318 CTGGAAGGCGAGAGCAACTTGAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1377
DB 1556 CTGGAAGGCGAGAGCAACTTGAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1615
QY 1378 AGGAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1437
DB 1616 AGGAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1675
QY 1438 GAGTTAGAGAGCTGTAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1497
DB 1676 GAGTTAGAGAGCTGTAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1735
QY 1498 TGTGCACTGAGCAACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1557
DB 1736 TGTGCACTGAGCAACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1795
QY 1558 ATTGCTGAATACCCACTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1617
DB 1796 ATTGCGGAATACCCACTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1855

```

```

QY 1618 CTTATTCAGAGAGAAACAGATCTCAGTACAGAGTAAACAGTCCAGAGAACAGTTTG 1677
DB 1856 CTTATTCAGAGAGAAACAGATCTCAGTACAGAGTAAACAGTCCAGAGAACAGTTTG 1915
QY 1678 CATAGAGACTGCTCTTCTTACCTTCAAAAGACCTTGGAGAGAGAGAGAGAGAGAG 1737
DB 1916 CACAGAGATTCACCTGTTTACACTTAAAGAGCCTTACAGAGAGAGAGAGAGAGAG 1975
QY 1738 CAGCTCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1797
DB 1976 CAGCTCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2035
QY 1798 GATGTTTCAACCAACCACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1857
DB 2036 GATGTTTCAACCAACCACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2095
QY 1858 AAGCAGAGGTCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1893
DB 2096 AAGCAGAGGTCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2131

```

```

RESULT 14
BD127640
LOCUS BD127640 1676 bp DNA linear PAT 18-SEP-2002
DEFINITION Primer for synthesizing full-length cDNA and use thereof.
ACCESSION BD127640
VERSION BD127640.1 GI:2322585
KEYWORDS JP 2002017375-A/3071.
SOURCE
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 1676)
Oca,T., Nishikawa,T., Isogai,T., Hayashi,K., Ishii,S., Kawai,Y.,
Wakamatsu,A., Sugiyama,T., Nagai,K., Kojima,S., Otsuki,T. and
Koga,H.

```

```

REFERENCE
AUTHORS Oca,T., Nishikawa,T., Isogai,T., Hayashi,K., Ishii,S., Kawai,Y.,
Wakamatsu,A., Sugiyama,T., Nagai,K., Kojima,S., Otsuki,T. and
Koga,H.
TITLE Primer for synthesizing full-length cDNA and use thereof
JOURNML Patent: JP 2002017375-A 3071 22-JAN-2002;
COMMENT HELIX RESEARCH INSTITUTE
OS Homo sapiens (human)
PN JP 2002017375-A/3071
PD 22-JAN-2002
PF 07-JUL-2000 JP 2000253172
PI TOSHIO OTA,TETSUO NISHIKAWA,TAKAO ISOgai,KOJI HAYASHI,SHIZUKO
PI ISHII
PI YURI KAMAI,AI WAKAMATSU,TOMOYASU SUGIYAMA,KEIICHI NAGAI, PI
SHINICHI KOJIMA,
PI TETSUJI OTSUKI,HISASHI KOGA
PC
C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/ PC
10', C12P21/02,C12Q1/68//C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC
Primer for synthesizing full-length cDNA and use thereof FH Key
Location/Qualifiers
FT CDS (264)..(1676).

```

```

FEATURES
source
1..1676
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

```

```

BASE COUNT 485 a 381 c 458 g 352 t
ORIGIN

```

```

Query Match 30.3%; Score 1104.2; DB 6; Length 1676;
Best Local Similarity 86.8%; Pred. No. 1.4e-239;
Matches 1227; Conservative 0; Mismatches 183; Indels 3; Gaps 1;

```

```

QY 1 ATGGCTAGTTTCCACACCTTGTGGTGTAGCTGATGTCGAGCATTAAGTGTGAG 60
DB 264 ATGGCTAGTTTCCACACCTTGTGGTGTAGCTGATGTCGAGCATTAAGTGTGAG 323
QY 61 GAAAGGCGCAGATGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120

```

```

Db 324 GAAAGCGAAGCATGATGACGAGTTTCATAGTTTAAAGCCAAATATCTGATTCATTACT 383
Qy 121 GGTGATCAAGCGAGAACTTTTTCATCTGGTGTACTGAGCTGTCTAGACAA 180
Db 384 GGTGATCAAGCTGAAACTTTTTTTTCAATCTGGTGTACTGAGCTGTCTAGACAA 443
Qy 181 ATATGGGCGCTAGCGACATGATTAAGATGAAAGATGATCAAGTGAATTTTTCATA 240
Db 444 ATATGGGCGCTAGCTGACATGATTAAGATGAAAGATGATCAAGTGAATTTTTCATA 503
Qy 241 GCGATGAGCTTATCAACTGAAAGCTACAGGATATGAGTCTCCCTCCACATCTCCCT 300
Db 504 GGTATGAACTTATCAACTGAAAGCTACAGGATATGAGTCTCCCTCCACATCTCCCT 563
Qy 301 GTCATGAAACAGCAACGAGTGTATTTCCAGTGCACGCAATTTGATAGAGGAGATT 360
Db 564 GTCATGAAACAGCAACGAGTGTATTTCCAGTGCACGCAATTTGATAGAGGAGATT 623
Qy 361 GCTAGCATGCCACCACTCAAGCTGTGCTGCTCCATGAGGCTCCATTCAGTTGTT 420
Db 624 GCGAGATGCCACCGCTTACAGCTGTGCTCCAGTGCATGAGGATTCATTCAGTTGTT 683
Qy 421 GGAATCTTCCACCTTATGATCTGTCTGCTCCAGCAGCAGTGTCTCCCTGCTTAAC 480
Db 684 GGAATGCTCCAACTTATGATCTGTCTGCTCCAGCAGCAGTGTCTCCCTGCTTAAC 743
Qy 481 GGGGCTCCCTCCGCTCAAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
Db 744 GGGGCTCCCTCCGCTTATTAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 803
Qy 541 AAGAGTCTCTCTTCCAGCAGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600
Db 804 AAGAGTCTCTCTTATGATGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 863
Qy 601 GCAATATCAATTCGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
Db 864 GCAATGCTCAATTCGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 923
Qy 661 TCAAGGCTGAATATCAAGGATATTCACAGCAGCAAGAACTTATGATGACATCTTA 720
Db 924 TCAAGGCTGAATATCAAGGATATTCACAGCAGCAAGAACTTATGATGACATCTTA 983
Qy 721 AAGAGTCTCTCTTCCAGCAGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780
Db 984 AAGAGTCTCTCTTATGATGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1043
Qy 781 TCAATATGGAATCTTCTGACATGATCAAGATGAAACTCACTGACAGAAATTTATC 840
Db 1044 TCAATATGGAATCTTCTGACATGATCAAGATGAAACTCACTGACAGAAATTTATC 1103
Qy 841 CTAGCTATGACATTAATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900
Db 1104 CTAGCTATGACATTAATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1163
Qy 901 CCAGAAATACATCCCTCTCTTCCAGAAAGTTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960
Db 1164 CCAGAAATACATCCCTCTCTTCCAGAAAGTTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1223
Qy 961 AGCTCTTCTCTGCTGATGAGAGCTGCTGAGAGCCGCTGCTGAGAGATGAGCAGCAG 1020
Db 1224 AGCTCTTCTCTGCTGATGAGAGCTGCTGAGAGCCGCTGCTGAGAGATGAGCAGCAG 1283
Qy 1021 C---CAGAGAAAGAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1077
Db 1284 CAATTAAGAAAGAACTTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1343
Qy 1078 GCGAGTGTGAGCTGAGAGAGCGCGCAGAGCGCTTTGAGAGCAGAGCGCAAGAGCAG 1137
Db 1344 GCGAGTGTGAGCTGAGAGAGCGCGCAGAGCGCTTTGAGAGCAGAGCGCGCAAGAGCAG 1403
Qy 1138 GAGCGGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1197
Db 1404 GAGCGGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1463

```

```

Qy 1198 CAGAGGCGCAAGCGGACGCTGAGCTGAGAGAGCAGCTGAGAGAGCAGCGGAGCTGAG 1257
Db 1464 CAGAGGCGCAAGCGGACGCTGAGCTGAGAGAGCAGCTGAGAGAGCAGCGGAGCTGAG 1523
Qy 1258 CCGCAGCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1317
Db 1524 CCGCAGCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1583
Qy 1318 CTGGAAGGCGCAGCAGCACTTGAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1377
Db 1584 CTGGAAGGCGCAGCAGCACTTGAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1643
Qy 1378 AGGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1410
Db 1644 AGGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1676

RESULT 15
AK075290
LOCUS 1676 bp mRNA linear PRI 03-SEP-2002
DEFINITION Homo sapiens CDNA FLJ90809 f16, clone Y79AA1000778, highly similar
to Homo sapiens interseclin long form mRNA.
ACCESSION AK075290
VERSION AK075290.1 GI:22761283
KEYWORDS oligo capping; f16 (full insert sequence).
SOURCE Homo sapiens
ORGANISM Homo sapiens (human)
REFERENCE 1. Isogai, T., Ota, T., Nishikawa, T., Hayashi, K., Otsuki, T., Sugiyama, T., Suzuki, Y., Nagai, K., Sugano, S., Ishii, S., Kawai-Hio, Y., Saito, K., Yamamoto, J., Wakamatsu, A., Nakamura, Y., Kojima, S., Nagahara, K., Masuho, Y., Oho, T., Okano, K., Yoshikawa, Y., Aotsuba, S., Sasaki, N., Hattori, A., Okumura, K., Iwayanagi, T. and Niimiya, K.
NEO human cDNA sequencing project
TITLE Unpublished
JOURNAL 2 (bases 1 to 1676)
REFERENCE Isogai, T. and Otsuki, T.
AUTHORS Direct Submission
TITLE Submitted (25-MAR-2002) Takao Isogai, Helix Research Institute, Genomics Laboratory, 1532-3 Yana, Kizarezu, Chiba 292-0812, Japan (E-mail: genomics@helix.co.jp, Tel: 81-438-52-3575, Fax: 81-438-52-3586)
COMMENT NEO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology; cDNA library construction: Institute of Medical Science, University of Tokyo; Laboratory of Genome Structure, Human Genome Center; cDNA 5'- & 3'-end one pass sequencing and clone selection: Helix Research Institute (supported by Japan Key Technology Center etc.).
FEATURES
source
1..1676
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="Y79AA1000778"
/cell_line="Y79"
/cell_type="retinoblastoma"
/clone_id="Y79AA1"
/note="Cloning vector: pME18FLJ3"
BASE COUNT 485 a 381 c 458 g 352 t
ORIGIN
Query Match 30.3%; Score 1104.2; DB 9; Length 1676;
Best Local Similarity 86.8%; Pred. No. 1.4e-239;
Matches 1227; Conservative 0; Mismatches 183; Indels 3; Gaps 1;
Qy 1 ATGGCTCAGTTCCACACCTTTGCGTGTAGCTGATGCTGTGGCCATACTGTGAG 60
Db 264 ATGGCTCAGTTCCACACCTTTTGTGTGAGAGCTGAGATCTGGCCATACTGTAGAG 323

```


Db 1621 GAACTCTGGAATCAGAGGAAACAAGGACAGGGGACCGTGCTCTGGAAGGCAAGAG 1680
Qy 1681 AAGACTGAGATTGAGTTAGTAGAGCTGGAATGACAAAAGCATCAGCTAGAGGAAA 1740
Db 1681 AAGACTGAGATTGAGTTAGTAGAGCTGGAATGACAAAAGCATCAGCTAGAGGAAA 1740
Qy 1741 CTTGAGATATCAGGATGTCGACTGGCAACCCAGAGGCAAGAAATTGAGAGCAAG 1800
Db 1741 CTTGAGATATCAGGATGTCGACTGGCAACCCAGAGGCAAGAAATTGAGAGCAAG 1800
Qy 1801 TTTAGAGACTAAGAAATGCTGAAATCACTTACAGACAGAGTTGAGAGAAATCTG 1860
Db 1801 TTTAGAGACTAAGAAATGCTGAAATCACTTACAGACAGAGTTGAGAGAAATCTG 1860
Qy 1861 CAATGCTTGAAGAATTATTCAGAGAAACAGATCTAGTGAACAGTTAAACAGATC 1920
Db 1861 CAATGCTTGAAGAATTATTCAGAGAAACAGATCTAGTGAACAGTTAAACAGATC 1920
Qy 1921 CAGCAGAACAGTTTGCATAGAGACTGCTTCTTACCTCAAAAGAGCTTGGAGCAAG 1980
Db 1921 CAGCAGAACAGTTTGCATAGAGACTGCTTCTTACCTCAAAAGAGCTTGGAGCAAG 1980
Qy 1981 GAGCTGGCCCGGACAGAGCTCCGGAGACGCTGACAGAGTGAAGAGAGACCAAGTCA 2040
Db 1981 GAGCTGGCCCGGACAGAGCTCCGGAGACGCTGACAGAGTGAAGAGAGACCAAGTCA 2040
Qy 2041 AAGCTGAGAGAAATTGATGTTTTCACAAACAGCTGAAGAGAACTGAGAGATATAC 2100
Db 2041 AAGCTGAGAGAAATTGATGTTTTCACAAACAGCTGAAGAGAACTGAGAGATATAC 2100
Qy 2101 AAACAGCAACTCCAGAAAGAGAGAGTCCCTGAGAGACGCGACTGAAGCAGAAAGAGCAG 2160
Db 2101 AAACAGCAACTCCAGAAAGAGAGAGTCCCTGAGAGACGCGACTGAAGCAGAAAGAGCAG 2160
Qy 2161 GAGAGAAAGGCTTGAGATTAGAGAGCAAGAAAGAAAGAGAGAGAGAGAGAGAGAG 2220
Db 2161 GAGAGAAAGGCTTGAGATTAGAGAGCAAGAAAGAAAGAGAGAGAGAGAGAGAGAG 2220
Qy 2221 AAGGCAAGGCAATGCTGAGAGATGTCGACGAGAGAGAGACCAACGCCCCGGAGAACCC 2280
Db 2221 AAGGCAAGGCAATGCTGAGAGATGTCGACGAGAGAGAGACCAACGCCCCGGAGAACCC 2280
Qy 2281 CACGAGAGAGACAGACTGAAGAGAGAGACAGTGTCAAGAGAAAGAGAGAGAGAGAG 2340
Db 2281 CACGAGAGAGACAGACTGAAGAGAGAGACAGTGTCAAGAGAAAGAGAGAGAGAGAG 2340
Qy 2341 GCCAAGCCGGAATGCAAGACAGAGAGAGTCCGCTTTTCATCCGATCAGAGAGCAGCT 2400
Db 2341 GCCAAGCCGGAATGCAAGACAGAGAGAGTCCGCTTTTCATCCGATCAGAGAGCAGCT 2400
Qy 2401 AAGCTGGGCAACCCAGAGCAACCTGCTGTACCAAGAGAAAGGCGCTTACATTTCTGCA 2460
Db 2401 AAGCTGGGCAACCCAGAGCAACCTGCTGTACCAAGAGAAAGGCGCTTACATTTCTGCA 2460
Qy 2461 CAGGAGAGTGAAGAGTGTATATTAACGAGGCTGTACCCCTTTGAATCCAGAGTCA 2520
Db 2461 CAGGAGAGTGAAGAGTGTATATTAACGAGGCTGTACCCCTTTGAATCCAGAGTCA 2520
Qy 2521 GATGAGATCACCATCAGCAGAGATATATGATCATGTGATGATAAGCCAGACTGAGAG 2580
Db 2521 GATGAGATCACCATCAGCAGAGATATATGATCATGTGATGATAAGCCAGACTGAGAG 2580
Qy 2581 CAGAGATGGCTTGGAGAGAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2640
Db 2581 CAGAGATGGCTTGGAGAGAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2640
Qy 2641 GAAAGATTCAGAGAAATAGAGTCCCATCCAGCCAAACAGTGAACGATCTGACATCT 2700
Db 2641 GAAAGATTCAGAGAAATAGAGTCCCATCCAGCCAAACAGTGAACGATCTGACATCT 2700
Qy 2701 GCCCTGCCCCCAACTGGCTCTGCGTGAAGACCCCTGCTCTTTGCAAGTGAACCTTCT 2760
Db 2701 GCCCTGCCCCCAACTGGCTCTGCGTGAAGACCCCTGCTCTTTGCAAGTGAACCTTCT 2760

Qy 2761 GAGCCCTCCACAACCCCAACAACCTGGAGAGATTCAGTTCCAGTGGGCCAGAGCTCA 2820
Db 2761 GAGCCCTCCACAACCCCAACAACCTGGAGAGATTCAGTTCCAGTGGGCCAGAGCTCA 2820
Qy 2821 AACGAGAACGAGAACGAGCAACTGGAGATAGTGGGCGGCTCAGCTTCTCTGAACCTGA 2880
Db 2821 AACGAGAACGAGAACGAGCAACTGGAGATAGTGGGCGGCTCAGCTTCTCTGAACCTGA 2880
Qy 2881 CTTAGTGTGGCAGATTACGGCAGATCAGCTTTACCCAGCCACAGCCACTGGCTCC 2940
Db 2881 CTTAGTGTGGCAGATTACGGCAGATCAGCTTTACCCAGCCACAGCCACTGGCTCC 2940
Qy 2941 TCCCATGCTCCGCTGAGGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3000
Db 2941 TCCCATGCTCCGCTGAGGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3000
Qy 3001 TATCCCTGAGAGCCCAAAAGAGCAACCACTTAATTTTAAACAAAGTGAACGTCAAC 3060
Db 3001 TATCCCTGAGAGCCCAAAAGAGCAACCACTTAATTTTAAACAAAGTGAACGTCAAC 3060
Qy 3061 GTTCTGAAACAGCAAGACATGTGTGTTGAGAGATTCAAGTCAAGAGGTTGTTTC 3120
Db 3061 GTTCTGAAACAGCAAGACATGTGTGTTGAGAGATTCAAGTCAAGAGGTTGTTTC 3120
Qy 3121 CCCAGTCTTACGTAAACTCATTTCAAGGCCCCGTAAGAGAAATCCACAAAGCATTCAT 3180
Db 3121 CCCAGTCTTACGTAAACTCATTTCAAGGCCCCGTAAGAGAAATCCACAAAGCATTCAT 3180
Qy 3181 GGCCTTACTGAAGTCTGCTAGTCTAAGAGAGAGGCTTCCCGGCGCCCAAGCAGCC 3240
Db 3181 GGCCTTACTGAAGTCTGCTAGTCTAAGAGAGAGGCTTCCCGGCGCCCAAGCAGCC 3240
Qy 3241 ATTCCCGAGAAAGATTATTCACATGTAACATACATACAGAGTTCGAGCAAGAGATT 3300
Db 3241 ATTCCCGAGAAAGATTATTCACATGTAACATACATACAGAGTTCGAGCAAGAGATT 3300
Qy 3301 ACCTTTCAAGAGAGAGAGATGATTTGTGTTTCCAGAGAAAGAGTGAAGTGGAGAG 3360
Db 3301 ACCTTTCAAGAGAGAGAGATGATTTGTGTTTCCAGAGAAAGAGTGAAGTGGAGAG 3360
Qy 3361 ACGGTGGGCGCAACAGTCCGAGAGTCTTCCCTTCTAATCTATGAGAGCTTAAAGATT 3420
Db 3361 ACGGTGGGCGCAACAGTCCGAGAGTCTTCCCTTCTAATCTATGAGAGCTTAAAGATT 3420
Qy 3421 GGCCTGGAACCTGCTGGAGAAACAGGAGATTAGAGAAACCTGAATTCAGAG 3480
Db 3421 GGCCTGGAACCTGCTGGAGAAACAGGAGATTAGAGAAACCTGAATTCAGAG 3480
Qy 3481 ATTGCTTCTTACGCTGCTACGTGTCCCGAACAACCTCAGCTGCTGGAGAGCTGAT 3540
Db 3481 ATTGCTTCTTACGCTGCTACGTGTCCCGAACAACCTCAGCTGCTGGAGAGCTGAT 3540
Qy 3541 CTGATCCGGAAGAAAGAAACCAAGTGAAGTGTGAGAGAGAGAACTCAAGCTGAGAG 3600
Db 3541 CTGATCCGGAAGAAAGAAACCAAGTGAAGTGTGAGAGAGAGAACTCAAGCTGAGAG 3600
Qy 3601 AAGGCCAGATAGGGTGTTCAGCAAAATATGCAACTTCAAGCCCGGAAACAGC 3660
Db 3601 AAGGCCAGATAGGGTGTTCAGCAAAATATGCAACTTCAAGCCCGGAAACAGC 3660
Qy 3661 AAAATCACCACCACTGACACTACCAAGACCCGAGTGAACGAGAGTGTGAGAGT 3720
Db 3661 AAAATCACCACCACTGACACTACCAAGACCCGAGTGAACGAGAGTGTGAGAGT 3720
Qy 3721 GGGATGTAAGATTACACCGCCAGAACGATGACGAACTTACAGCAAGAGGCGAGATC 3780
Db 3721 GGGATGTAAGATTACACCGCCAGAACGATGACGAACTTACAGCAAGAGGCGAGATC 3780
Qy 3781 ATCAAGCTCTCAACAAAGAGAGACCCGAGCTGTGTGAAGAGAGAGTCAAGTGGCA 3840
Db 3781 ATCAAGCTCTCAACAAAGAGAGACCCGAGCTGTGTGAAGAGAGAGTCAAGTGGCA 3840

QY	3841	GGGCGCTTCCCATCCCAATTAAGTAAGCGACACAGACATGGACCCGAGCAGAATGA	3900
Db	3840	GGGCTCTTCCCATCCCAATTAATGTAAGCTGACACAGACATGGACCCGAGCAGAATGA	3899
QY	3901	ATCATATGTGTGTCATATCCCGCCCTCAGAGCTTGAAAGTCCCTCAAGAGACCCACTATCCCA	3966
Db	3900	ATCATATGTGTGTCCATATCCCGCCCTCAGAGCTTGAAAGTCTCAAGAGACCCACTATCCCA	3959
QY	3961	TATCATCTGCCAGAGGGATGATGGAGATGACAGCTTGATCATGTGACTTGACATGAT	4020
Db	3960	TATCATCTGCCAGAGGGATGATGGAGATGACAGCTTGATCATGTGACTTGACATGAT	4019
QY	4021	CACCTACTGCCCTTCGATGAGAGAAACTCACGACGACAGTTTACCTCATTTGACCTTA	4080
Db	4020	CACCTACTGCCCTTCGATGAGAGAAACTCACGACGACAGTTTACCTCATTTGACCTTA	4079
QY	4081	GTTGCATGTGATCGAAATGTCTGATGTCACTGCGTCCAGAGGACAGAGCAAAATTGCAGAAC	4140
Db	4080	GTTGCATGTGATCGAAATGTCTGATGTCACTGCGTCCAGAGGACAGAGCAAAATTGCAGAAC	4139
QY	4141	TGCACAGGGTGTGGGTCTTTTGGGGCTTTCCTAGTCACTCAGACTGACCGGCCCGCC	4200
Db	4140	TGCACAGGGTGTGGGTCTTTTGGGGCTTTCCTAGTCACTCAGACTGACCGGCCCGCC	4199
QY	4201	TTTCACACGGGGCCTTTCATATCTTTTAAATATATTTTAAATGCTATTTTAACTCTTTTA	4266
Db	4200	TTTCACACGGGGCCTTTCATATCTTTTAAATATATTTTAAATGCTATTTTAACTCTTTTA	4259
QY	4261	ATTAATAATCTCAATCAATTACTCTTTCGCTATTTTGGTCTTACAAATAACCCCACTATC	4320
Db	4260	ATTAATAATCTCAATCAATTACTCTTTCGCTATTTTGGTCTTACAAATAACCCCACTATC	4319
QY	4321	AAAGAGTGTCTGTCTGCGGACGATTAAATGTCTGTTCCGGGGGTACCGTAACTGAGAGC	4380
Db	4320	AAAGAGTGTGTCTGTCTGCGGACGATTAAATGTCTGTTCCGGGGGTACCGTAACTGAGAGC	4379
QY	4381	TTGCTGTATCTTGCGCTTGTGTCAGAGTGTCCCAACCACTGTGTATGTTGGGCTGT	4440
Db	4380	TTGCTGTATCTTGCGCTTGTGTCAGAGTGTCCCAACCACTGTGTATGTTGGGCTGT	4439
QY	4441	CCCTGCCGTAGAGCAACAGAGAGATGGGGTACTGTTTGAATAATGTATGTAGACTG	4500
Db	4440	CCCTGCCGTAGAGCAACAGAGAGATGGGGTACTGTTTGAATAATGTATGTAGACTG	4499
QY	4501	AGCTGTACTTGAAGGGGGTTATGCTGTCTGTGACCACTACGTTACTGTCCGCGCATG	4560
Db	4500	AGCTGTACTTGAAGGGGGTTATGCTGTCTGTGACCACTACGTTACTGTCCGCGCATG	4559
QY	4561	TACCATCTGTACCGAAGAAATGCTCTTCTCCATAGCGTAAACCAACACCGGTGTACGT	4622
Db	4560	TACCATCTGTACCGAAGAAATGCTCTTCTCCATAGCGTAAACCAACACCGGTGTACGT	4619
QY	4621	GCTTCATCTTACCTGAATTCATTTTACTTTTGACACAGTACCTTTGTAGCCA-CTGAGAAAGC	4680
Db	4620	GCTTCATCTTACCTGAATTCATTTTACTTTTGACACAGTACCTTTGTAGCCA-CTGAGAAAGC	4679
QY	4681	ACCCTATGTTCCGTTTGTGCTCAGATGTACCTTATGTTGCCCCGTGTTTGTATTTT	4740
Db	4679	ACCCTATGTTCCGTTTGTGCTCAGATGTACCTTATGTTGCCCCGTGTTTGTATTTT	4739
QY	4741	TCAATCTGGCAGTGTTCACACCATTAACCTAGTAAGACGCCAACCTGCCAGCGGTACG	4800
Db	4739	TCAATCTGGCAGTGTTCACACCATTAACCTAGTAAGACGCCAACCTGCCAGCGGTACG	4798
QY	4801	ATCATCAGTACCCACCGTCTTATGTCTCTGTACGTAAGTATTTATTCAGTTGCTTTTAT	4866
Db	4799	ATCATCAGTACCCACCGTCTTATGTCTCTGTACGTAAGTATTTATTCAGTTGCTTTTAT	4855
QY	4861	GGATATCTTTGAACAAGTATCTTCTTGAACAAGAAATGTATAGAAAGTCTCCCTGCAG	4920
Db	4859	GGATATCTTTGAACAAGTATCTTCTTGAACAAGAAATGTATAGAAAGTCTCCCTGCAG	4919
QY	4921	TTTAAATTTCCCAAGTATTACATTTTAACTGACGTGGGGGTGCTACATTAATATG	4980

Db	4919	TTAATTTCCAGTGTGTTACATTTTTTAACTAGACGTGGGGTTCTCAAGTTAATAAG	4972
Qy	4961	AAATGGCGCTCCTGCTCCGTGTGTGTGTTAACTTGCTGTAGCTGAAGCCGTGTCTCT	5044
Db	4979	AAATGGCGCTCCTGGATCGGTGTGTGTGTTAACTTGCTGTGTAGCTGAAGCCGTGTCTCT	5033
Qy	5041	TAGATATTAGTTGGAGAGTCGGGAAGAGAAATTCGATATCAAGTT	5084
Db	5039	TAGATATTAGTTGGAGAGTCGGGAAGAGAAATTCGATATCAAGTT	5082
RESULT 2			
ID	AAZ39024	standard; cDNA, 5738 BP.	
XX	AAZ39024;		
AC	28-FEB-2000	(first entry)	
XX			
DT			
XX			
DE		Mouse Ese1l cDNA sequence.	
XX			
XX		Mouse; murine; Ese1; Ese2; endocytosis; vesicular trafficking;	
KW		regulation; actin cytoskeleton; detection; cancer; infection;	
KW		EH-domain and SH3-domain regulator of endocytosis; anticancer;	
KM		antiproliferative; antiviral; ss.	
XX			
OS		Mus sp.	
XX			
PN		WO955728-A2.	
XX			
PD		04-NOV-1999.	
XX			
PF		27-APR-1999; 99WO-CA00375.	
XX			
PR		27-APR-1998; 98CA-2230201.	
PR		05-FEB-1999; 99US-0118739.	
PA		(HSCR-) HSC RES & DEV LP.	
XX			
XX		Egan SE, Wang W, Sengar A;	
PI		WPI, 2000-052802/04.	
DR		P-PSDB; AA57449.	
XX			
PT		New nucleic acid encoding Ese1 and 2 proteins, involved in regulation	
PT		of endocytosis, used e.g. for treating cancer or preventing viral	
PT		infection _	
XX			
PS		Claim 6, Page 56-59; 99pp; English.	
XX			
CC		The present invention specifically describes mammalian Ese1 and 2	
CC		proteins (I) and their splice variants (Ese = EH-domain and SH3-domain	
CC		regulator of endocytosis). (II) are involved in regulation of clathrin-	
CC		mediated endocytosis (as a complex with Esp1s protein), vesicular	
CC		trafficking and actin cytoskeleton. Generally (I) (or its (ant)agonists,	
CC		mimetics, fragments and inactive mutants); (I)-specific antibodies (Ab);	
CC		sequences antisense to the (I) polynucleotide; agents that downregulate	
CC		expression of Ese genes or antagonists of an Ese binding partner are	
CC		used to treat diseases associated with undesirable endocytosis and	
CC		resulting changes in cellular function. Particularly overexpression of	
CC		Ese1 is used to block clathrin-mediated endocytosis in vivo or in cell	
CC		cultures, while administration of (I) is used to promote endocytosis of	
CC		selected cells. (Ant)agonists of (I) or Ab are used to suppress abnormal	
CC		proliferation of cells that can be stimulated to proliferate by a growth	
CC		factor receptor; and similar compounds (also inactive Ese mutants) can be	
CC		used to prevent viral infection. Endocytosis may also be regulated, in	
CC		vivo or in cell cultures, by forming an Ese-Esp1s complex, then binding	
CC		dynamin to the complex. Generally conditions that can be treated include	
CC		cancer; abnormal cell division or migration; viral infection; or abnormal	
CC		receptor signalling, tissue development or synaptic transmission. The	
CC		present sequence represents mouse Ese1l cDNA sequence.	
XX			

SQ Sequence 5738 BP; 1556 A; 1456 C; 1613 G; 1113 T; 0 other;

Query Match 76.4%; Score 3886.4; DB 21; Length 5738;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 3898; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

1 CGGCAAGAGAGAGTGGAGCCGCGC - GGGAGGCGCGCAGCTGGTTCCTCCAGTAC 59
1 CGGCAAGAGAGAGTGGAGCCGCGCGGGAGGGCGCGCAGCTGGTTCCTCCAGTAC 60
60 GCGCGCTGGCAAGAGAGCATCCCGAGCGGGCTCCGGGACCGCGCGAGGCAAGCGCG 119
61 GCGCGCTGGCAAGAGAGCATCCCGAGCGGGCTCCGGGACCGCGCGAGGCAAGCGCG 120
120 GCGCGCGGGAGATGATGCGCGCGCTGCGGACCTCGGCTTCGCGCGCGCGCGCGCGCTG 179
121 GCGCGCGGGAGATGATGCGCGCGCTGCGGACCTCGGCTTCGCGCGCGCGCGCGCGCTG 180
180 CACTGATTTGTGTGAGGGCGCGCGCGCGCGCGCGCGCGCGAGATGAGCGCTGATCAGA 239
181 CACTGATTTGTGTGAGGGCGCGCGCGCGCGCGCGCGCGCGAGATGAGCGCTGATCAGA 240
240 AGGTAAAGCTAAATAGAACCATGCTCAGTTTCCACACCTTCGCTGAGCTGAGT 299
241 AGGTAAAGCTAAATAGAACCATGCTCAGTTTCCACACCTTCGCTGAGCTGAGT 300
300 CTGGGCGCTAACTGTGAGGAGAAAGGGCGCAGCATGACCGAGTTCCTTACCTGAAGC 359
301 CTGGGCGCTAACTGTGAGGAGAAAGGGCGCAGCATGACCGAGTTCCTTACCTGAAGC 360
360 GATAGCGGAGTTTATTACTGTGATCAAGCGAGAACTTTTTCCTTCCATCTGGGTACC 419
361 GATAGCGGAGTTTATTACTGTGATCAAGCGAGAACTTTTTCCTTCCATCTGGGTACC 420
420 TCAGCTCTGCTTAGACAACAATATGCGCGCTAGCGGACATGAATTAAGATGAAGATGA 479
421 TCAGCTCTGCTTAGACAACAATATGCGCGCTAGCGGACATGAATTAAGATGAAGATGA 480
480 TCAATTTGAAATTTTTCATAGCCATGAAGTTTATCAAACTGAAGCTTCAAGATATCAGT 539
481 TCAATTTGAAATTTTTCATAGCCATGAAGTTTATCAAACTGAAGCTTCAAGATATCAGT 540
540 CCCCTCCACACTCCCTCCCTCATGAAACAGCAACAGTGGCTATTTCCAGTGCACAC 599
541 CCCCTCCACACTCCCTCCCTCATGAAACAGCAACAGTGGCTATTTCCAGTGCACAC 600
600 ATTGGAATAGGAGGAGTTGCTAGCAATGCGACCACTCAGCTGTGCTCTGTGCAAT 659
601 ATTGGAATAGGAGGAGTTGCTAGCAATGCGACCACTCAGCTGTGCTCTGTGCAAT 660
660 GGGCTTCATTCAGTTTGGAAATGCTCACCCTTGAATCTTCTGCTCCCTCCAGCAGC 719
661 GGGCTTCATTCAGTTTGGAAATGCTCACCCTTGAATCTTCTGCTCCCTCCAGCAGC 720
720 AGTGCCTCCCTGGCTAACGGGGCTCCTCCGTCATACAGCCTGCGCTGGCTTGGCGA 779
721 AGTGCCTCCCTGGCTAACGGGGCTCCTCCGTCATACAGCCTGCGCTGGCTTGGCGA 780
780 TCCTCAGGACATGCGCCAAAGAGTCTTCTTCAGCAGATCTGCTCAGGCTCAAT 839
781 TCCTCAGGACATGCGCCAAAGAGTCTTCTTCAGCAGATCTGCTCAGGCTCAAT 840
840 AAACTAAGTTTACAGAAAGGCAATCATTCATGTCGACGCGCTCCAGCAGCAGA 899
841 AAACTAAGTTTACAGAAAGGCAATCATTCATGTCGACGCGCTCCAGCAGCAGA 900
900 ATGGGCTGCTCAGTCAAGGCTGAATATACAGGAGTATTCACAGCAGCAGCA 959
901 ATGGGCTGCTCAGTCAAGGCTGAATATACAGGAGTATTCACAGCAGCAGCA 960
960 AACTATAGATGAGCACTTAACAGGTCCTCCAGGCAAGAACTATTCATGCAATCAAGTTT 1019
961 AACTATAGATGAGCACTTAACAGGTCCTCCAGGCAAGAACTATTCATGCAATCAAGTTT 1020

1020 ACCCGAGCTCAGCTGCTTCAATATGGAATCTTCGACATTTGATCAAGATGAAAACT 1079
1021 ACCCGAGCTCAGCTGCTTCAATATGGAATCTTCGACATTTGATCAAGATGAAAACT 1080
1080 CACTCAGAGAAATTTATCTAGCTATGACCTTAATATGTTTGCATGCTGTCAGCC 1139
1081 CACTCAGAGAAATTTATCTAGCTATGACCTTAATATGTTTGCATGCTGTCAGCC 1140
1140 ACTGCGCGCGCTGCTGCTTCCAGAAATACATCCCTCTTCCAGAAAGTTCGCTCCGG 1199
1141 ACTGCGCGCGCTGCTGCTTCCAGAAATACATCCCTCTTCCAGAAAGTTCGCTCCGG 1200
1200 CAGTGGAGATGCTGCTGCTTCCAGAAATACATCCCTCTTCCAGAAAGTTCGCTCCGG 1259
1201 CAGTGGAGATGCTGCTGCTTCCAGAAATACATCCCTCTTCCAGAAAGTTCGCTCCGG 1260
1260 GTCAAGAGATGAGCAGCAGCAGCAGAACTGCTGTGACATTTGAAGATTAAGAGCG 1319
1261 GTCAAGAGATGAGCAGCAGCAGCAGAACTGCTGTGACATTTGAAGATTAAGAGCG 1320
1320 GAGAACTTTCAGAGCAGGAGCTGAGCTGAGAGAGCGCGCAAGCGCTCTTGGAGCA 1379
1321 GAGAACTTTCAGAGCAGGAGCTGAGCTGAGAGAGCGCGCAAGCGCTCTTGGAGCA 1380
1380 GAGAGCTTTCAGAGCAGGAGCTGAGCTGAGAGAGCGCGCAAGCGCTCTTGGAGCA 1439
1381 GAGAGCTTTCAGAGCAGGAGCTGAGCTGAGAGAGCGCGCAAGCGCTCTTGGAGCA 1440
1440 GCGGAGCGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1499
1441 GCGGAGCGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1500
1500 GAGCGGAGCTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1559
1501 GAGCGGAGCTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1560
1560 GCGCGGAGCTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1619
1561 GCGCGGAGCTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1620
1620 GGAATCTCTGATCAGAGGAAACAAGAGCAGAGGAGCAGGCTGCTGAGAGGCAAGAG 1679
1621 GGAATCTCTGATCAGAGGAAACAAGAGCAGAGGAGCAGGCTGCTGAGAGGCAAGAG 1680
1681 GAAAGCTCTGAGGTTTGAATAGAACTCTGATGACAAAAGCATCAGCTGAGAGGAAA 1740
1740 ACTTGAAGATATCAGGTCGACTGAGCAACCAAGGCAAGAAATTTGAGAGCAACAA 1799
1741 ACTTGAAGATATCAGGTCGACTGAGCAACCAAGGCAAGAAATTTGAGAGCAACAA 1800
1800 GTCTAGAGAGCTAAGATTTGCTGAATATCACCACCTTACAGCAGCAGTTCAGAGAAATCTCA 1859
1801 GTCTAGAGAGCTAAGATTTGCTGAATATCACCACCTTACAGCAGCAGTTCAGAGAAATCTCA 1860
1860 GCAAAATGCTTGAAGACTTATTCAGAGAAACAATACATCAGTGCAGATTAAAACAAGT 1919
1861 GCAAAATGCTTGAAGACTTATTCAGAGAAACAATACATCAGTGCAGATTAAAACAAGT 1920
1920 CAGAGCAAAAGTTTGCATAGAGACTCGCTTCTTACCTCAAAAGAGCTTGAAGCAAA 1979
1921 CAGAGCAAAAGTTTGCATAGAGACTCGCTTCTTACCTCAAAAGAGCTTGAAGCAAA 1980
1980 GAGAGCTGCGCGCAGCAGCTCCGGAGCAGCTGAGCAGGCTGAGAGAGAGCAGAGGTC 2039
1981 GAGAGCTGCGCGCAGCAGCTCCGGAGCAGCTGAGCAGGCTGAGAGAGAGCAGAGGTC 2040
2040 AAGCTGAGAGAGATGATGTTTTCACAAACAGCTGAAGAACTGAGAGAGATACATAG 2099
2041 AAGCTGAGAGAGATGATGTTTTCACAAACAGCTGAAGAACTGAGAGAGATACATAG 2100

```

QY 2100 CAACAGCACTCCAGAGAGAGGTCCTGAGGAGCAGCGCACTGAAGCAGAAAGACA 2159
DB 2101 CAACAGCACTCCAGAGAGAGGTCCTGAGGAGCAGCGCACTGAAGCAGAAAGACA 2160
QY 2160 GGAAGAGAGAGGTCGAGGTTAGAGAGCAAAAGAGAGAGGTCGAGAGAGAGTTCAAGA 2219
DB 2161 GGAAGAGAGAGGTCGAGGTTAGAGAGCAAAAGAGAGAGGTCGAGAGAGAGTTCAAGA 2220
QY 2220 AAGGAGAGAGAGGTCGAGGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2279
DB 2221 AAGGAGAGAGAGGTCGAGGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2280
QY 2280 CCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2339
DB 2281 CCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2340
QY 2340 AGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2399
DB 2341 AGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2400
QY 2400 TAAAGTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2459
DB 2401 TAAAGTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2460
QY 2460 ACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2519
DB 2461 ACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2520
QY 2520 CGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2579
DB 2521 CGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2580
QY 2580 GCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2639
DB 2581 GCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2640
QY 2640 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2699
DB 2641 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2700
QY 2700 TGCCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2759
DB 2701 TGCCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2760
QY 2760 TGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2819
DB 2761 TGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2820
QY 2820 AAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2879
DB 2821 AAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2880
QY 2880 ACCTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2939
DB 2881 ACCTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2940
QY 2940 CTCCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2999
DB 2941 CTCCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3000
QY 3000 GTATCCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3059
DB 3001 GTATCCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3060
QY 3060 CGTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3119
DB 3061 CGTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3120
QY 3120 CCCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3179
DB 3121 CCCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3180
QY 3180 TGGGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3239

```

```

DB 3181 TGGGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3240
QY 3240 CATTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3299
DB 3241 CATTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3300
QY 3300 AACCTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3359
DB 3301 AACCTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3360
QY 3360 AACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3419
DB 3361 AACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3420
QY 3420 GGGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3479
DB 3421 GGGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3480
QY 3480 TATTGCTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3539
DB 3481 TATTGCTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3540
QY 3540 TCTGATCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3599
DB 3541 TCTGATCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3600
QY 3600 AAAGCGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3659
DB 3601 AAAGCGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3660
QY 3660 CAATAATCAACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3719
DB 3661 CAATAATCAACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3720
QY 3720 CGGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3779
DB 3721 CGGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3780
QY 3780 CATCAACGTCCTCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3839
DB 3781 CATCAACGTCCTCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3840
QY 3840 TGGGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3899
DB 3841 TGGGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3900

RESULT 3
AAZ39009
ID AAZ39009 standard; cDNA; 3723 BP.
XX
AC AAZ39009;
XX
AC
XX
DT 28-FEB-2000 (first entry)
XX
DE Mouse BseI coding sequence.
XX
KW Mouse; murine; BseI; Bse2; endocytosis; vesicular trafficking;
KW regulation; actin cytoskeleton; detection; cancer; infection;
KW EH-domain and SH3-domain regulator of endocytosis; anticancer;
KW antiproliferative; antiviral; ss.
XX
OS Mus sp.
XX
PN W09955728-A2.
XX
PD 04-NOV-1999.
XX
PF 27-APR-1999; 99WO-CA00375.
XX
PR 27-APR-1998; 98CA-2230201.
PR 05-FEB-1999; 99US-0118739.

```

XX (HSCR-) HSC RES & DEV LP.
 XX Egan SE, Wang W, Sengar A;
 XX WPI: 2000-052802/04.
 XX P-PSDB; AAY57444.
 DR New nucleic acid encoding Ese1 and 2 proteins, involved in regulation
 PT of endocytosis, used e.g. for treating cancer or preventing viral
 PT infection -
 PS Claim 6: Page 40-42; 99pp; English.

XX The present sequence encodes mouse Ese1. The present invention
 CC specifically describes mammalian Ese1 and 2 proteins (I) and their splice
 CC variants (Ise = Esi-domain and Shi3-domain regulator of endocytosis). (I)
 CC are involved in regulation of clathrin-mediated endocytosis (as a complex
 CC with Eps15 protein), vesicular trafficking and actin cytoskeleton.
 CC Generally (I) (or its (ant)agonists, mimetics, fragments and inactive
 CC mutants); (I)-specific antibodies (Ab); sequences antisense to the (I)
 CC polynucleotide; agents that downregulate expression of Ese genes or
 CC antagonists of an Ese binding partner are used to treat diseases
 CC associated with undesirable overexpression of Ese1 is used to block
 CC function. Particularly overexpression of Ese1 is used to block
 CC clathrin-mediated endocytosis in vivo or in cell cultures, while
 CC administration of (I) is used to promote endocytosis of selected cells.
 CC (Ant)agonists of (I) or Ab are used to suppress abnormal proliferation of
 CC cells that can be stimulated to proliferate by a growth factor receptor;
 CC and similar compounds (also inactive Ese mutants) can be used to prevent
 CC viral infection. Endocytosis may also be regulated, in vivo or in cell
 CC cultures, by forming an Ese-Eps15 complex, then binding dynamin to the
 CC complex. Generally conditions that can be treated include cancer;
 CC abnormal cell division or migration; viral infection; or abnormal
 CC receptor signalling, tissue development or synaptic transmission.
 XX Sequence 3723 BP; 1065 A; 941 C; 1025 G; 692 T; 0 other;

Query Match 73.2%; Score 3723; DB 21; Length 3723;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 3723; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 259 ATGGCTCAATTTCCACACCTTTCGCTGCTAGCTGATGCTGGCCATTAATCTGAG 318
 DB 1 ATGGCTCAATTTCCACACCTTTCGCTGCTAGCTGATGCTGGCCATTAATCTGAG 60
 QY 319 GAAAGGGCCAGACATGACGAGCACTTCTAGGCTGAGCCGATGCGGATTTTATTCT 378
 DB 61 GAAAGGGCCAGACATGACGAGCACTTCTAGGCTGAGCCGATGCGGATTTTATTCT 120
 QY 379 GGTGATCAAGCGAGAACTTTTTCATCTGGGTTCCTCAGCCTTCTTACGACA 438
 DB 121 GGTGATCAAGCGAGAACTTTTTCATCTGGGTTCCTCAGCCTTCTTACGACA 180
 QY 439 ATATGAGGCGCTAGCGGACATGATTAAGTGAAGATGATCAAGTGAATTTTCCATA 498
 DB 181 ATATGAGGCGCTAGCGGACATGATTAAGTGAAGATGATCAAGTGAATTTTCCATA 240
 QY 499 GCCATGAAGCTTATCAAACTGAAGCTACAGAGATCACTCCCTCCACACTTCCCTC 558
 DB 241 GCCATGAAGCTTATCAAACTGAAGCTACAGAGATCACTCCCTCCACACTTCCCTC 300
 QY 559 GTCAAGAAACAGACACAGTGGCTATTTCCAGTGCACGCAATTTGATTAAGAGGAT 618
 DB 301 GTCAAGAAACAGACACAGTGGCTATTTCCAGTGCACGCAATTTGATTAAGAGGAT 360
 QY 619 GGTAGATGCACTCACTCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 678
 DB 361 GGTAGATGCACTCACTCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 420
 QY 679 GGAATGCTCCACCTTATGATCTTCTGCTCCCTCCAGACAGAGTGCCTCCCTGCTAAC 738
 DB 421 GGAATGCTCCACCTTATGATCTTCTGCTCCCTCCAGACAGAGTGCCTCCCTGCTAAC 480

QY 739 GGGGCTCCTCCGTCATACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 798
 DB 481 GGGGCTCCTCCGTCATACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 540
 QY 799 AAGAGTTCTTCTTCCAGAGATCTGATCCAGGGTCACATTAAGAATAAGTAAAGAG 858
 DB 541 AAGAGTTCTTCTTCCAGAGATCTGATCCAGGGTCACATTAAGAATAAGTAAAGAG 600
 QY 859 GCAATATCTTGAATGTCGACGCGCCCTCCAGACAGAGATGAGCTGCTGCTGCTGCT 918
 DB 601 GCAATATCTTGAATGTCGACGCGCCCTCCAGACAGAGATGAGCTGCTGCTGCTGCT 660
 QY 919 TCAAGGCTGAATTAAGAGGATTAATCAACGCGACGCAAACTATGATGAGCACTTA 978
 DB 661 TCAAGGCTGAATTAAGAGGATTAATCAACGCGACGCAAACTATGATGAGCACTTA 720
 QY 979 ACAGGTCCCAAGGCAAGAACTATTCATGCAATCAAGTTTACCCAGGCTCAGCTGCT 1038
 DB 721 ACAGGTCCCAAGGCAAGAACTATTCATGCAATCAAGTTTACCCAGGCTCAGCTGCT 780
 QY 1039 TCAATATGGAATCTTTCTGACATTTGATCAATGAGAAATCTCATGCAAGAAATTTATC 1098
 DB 781 TCAATATGGAATCTTTCTGACATTTGATCAATGAGAAATCTCATGCAAGAAATTTATC 840
 QY 1099 CTAGCTATGCACTTAATGATGTTGTCATGCTGCTGCTGCTGCTGCTGCTGCTGCT 1158
 DB 841 CTAGCTATGCACTTAATGATGTTGTCATGCTGCTGCTGCTGCTGCTGCTGCTGCT 900
 QY 1159 CCAGAAATCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1218
 DB 901 CCAGAAATCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 960
 QY 1219 AGCTCTTCTCTGCTGATCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1278
 DB 961 AGCTCTTCTCTGCTGATCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020
 QY 1279 CCAGAAAGAAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1338
 DB 1021 CCAGAAAGAAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1080
 QY 1339 AGTGTGAGCTGGAAGAGCGCCGCAAGCGCTCTTGAAGCAGACGCAAGAGAGAG 1398
 DB 1081 AGTGTGAGCTGGAAGAGCGCCGCAAGCGCTCTTGAAGCAGACGCAAGAGAGAG 1140
 QY 1399 CGTTGGCTCAGCTGAGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1458
 DB 1141 CGTTGGCTCAGCTGAGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
 QY 1459 GAGGCGAAGCGGACAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1518
 DB 1201 GAGGCGAAGCGGACAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
 QY 1519 CAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1578
 DB 1261 CAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
 QY 1579 GAAAGCGACGCAAACTTGAATGGAACGAAACCGAGAGACAGAACTCTGTAATCAGAG 1638
 DB 1321 GAAAGCGACGCAAACTTGAATGGAACGAAACCGAGAGACAGAACTCTGTAATCAGAG 1380
 QY 1639 AACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1698
 DB 1381 AACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
 QY 1699 TTAGAAGCTCTGAATGACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1758
 DB 1441 TTAGAAGCTCTGAATGACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500
 QY 1759 CGACTGGCAACCCAGAGGCAAGAAATTAAGAGACAGAACTTGAAGAGCTTAAGAAAT 1818
 DB 1501 CGACTGGCAACCCAGAGGCAAGAAATTAAGAGACAGAACTTGAAGAGCTTAAGAAAT 1560

QY 1819 GCTGAATACCCCACTTACAGACAGAGTTCAGAAATCTCAGAAATGCTTGAAGATT 1878
|
|
|
Db 1861 GCTGAATACCCCACTTACAGACAGAGTTCAGAAATCTCAGAAATGCTTGAAGATT 1820
|
|
|
QY 1879 ATTCAGAGAAACAGATATCTAGTGAACAGTTAAACAGTTCAGACAGAAAGTTTGAT 1938
|
|
|
Db 1821 ATTCAGAGAAACAGATATCTAGTGAACAGTTAAACAGTTCAGACAGAAAGTTTGAT 1880
|
|
|
QY 1939 AGAGATGCTTTTAAACCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1998
|
|
|
Db 1861 AGAGATGCTTTTAAACCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1740
|
|
|
QY 1999 CTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2058
|
|
|
Db 1741 CTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1800
|
|
|
QY 2059 GTTTTCAACAAACAGCTGAAGAACTGAGAGATATCATAGCAACAGCAATCTCCAGAG 2118
|
|
|
Db 1801 GTTTTCAACAAACAGCTGAAGAACTGAGAGATATCATAGCAACAGCAATCTCCAGAG 1860
|
|
|
QY 2119 CAGAGTCCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2178
|
|
|
Db 1861 CAGAGTCCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1920
|
|
|
QY 2179 TTAGAGAGCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2238
|
|
|
Db 1921 TTAGAGAGCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1980
|
|
|
QY 2239 GAGCATGTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2298
|
|
|
Db 1981 GAGCATGTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2040
|
|
|
QY 2299 AAGAGGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2358
|
|
|
Db 2041 AAGAGGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2100
|
|
|
QY 2359 GACAAGCAGAGTCCGCTTTTCCATCCGATCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2418
|
|
|
Db 2101 GACAAGCAGAGTCCGCTTTTCCATCCGATCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2160
|
|
|
QY 2419 CCTGTGTACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2478
|
|
|
Db 2161 CCTGTGTACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2220
|
|
|
QY 2479 GTATATTTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2538
|
|
|
Db 2221 GTATATTTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2280
|
|
|
QY 2539 CCAGAGATATATGTCATGATGATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2598
|
|
|
Db 2281 CCAGAGATATATGTCATGATGATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2340
|
|
|
QY 2599 GAGCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2658
|
|
|
Db 2341 GAGCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2400
|
|
|
QY 2659 GAGGTTCCTCACTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2718
|
|
|
Db 2401 GAGGTTCCTCACTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2460
|
|
|
QY 2719 GCTCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2778
|
|
|
Db 2461 GCTCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2520
|
|
|
QY 2779 AACAACTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2838
|
|
|
Db 2521 AACAACTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2580
|
|
|
QY 2839 GACAACATGGAGATATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2898
|
|
|
Db 2581 GACAACATGGAGATATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2640
|
|
|
QY 2899 CGGAGAGATAGAGCTTTTACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2958
|
|
|

Db 2641 CGGAGAGATAGAGCTTTTACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2700
|
|
|
QY 2959 GGGCAGGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3018
|
|
|
Db 2701 GGGCAGGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2760
|
|
|
QY 3019 AAAAGACAACACTTAAATTTTAAACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3078
|
|
|
Db 2761 AAAAGACAACACTTAAATTTTAAACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2820
|
|
|
QY 3079 ATGTGTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3138
|
|
|
Db 2821 ATGTGTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2880
|
|
|
QY 3139 CTCATTTTCAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3198
|
|
|
Db 2881 CTCATTTTCAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2940
|
|
|
QY 3199 GCTAGTCTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3258
|
|
|
Db 2941 GCTAGTCTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3000
|
|
|
QY 3259 ATTCGACATGATACATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3318
|
|
|
Db 3001 ATTCGACATGATACATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3060
|
|
|
QY 3319 GTGATTTGTGTTTACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3378
|
|
|
Db 3061 GTGATTTGTGTTTACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3120
|
|
|
QY 3379 GGAATCTTCCCTTAACTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3438
|
|
|
Db 3121 GGAATCTTCCCTTAACTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3180
|
|
|
QY 3439 AAAACAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3498
|
|
|
Db 3181 AAAACAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3240
|
|
|
QY 3499 ACTGTGTCCGAAACAATCACCCTGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3558
|
|
|
Db 3241 ACTGTGTCCGAAACAATCACCCTGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3300
|
|
|
QY 3559 CCAGAGTGAATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3618
|
|
|
Db 3301 CCAGAGTGAATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3360
|
|
|
QY 3619 TTTTCAGCAAAATTAATGTCAAATCTTAAGCCCGGAAACAAGCAAAATCACCCCAATGAG 3678
|
|
|
Db 3361 TTTTCAGCAAAATTAATGTCAAATCTTAAGCCCGGAAACAAGCAAAATCACCCCAATGAG 3420
|
|
|
QY 3679 CTACCCAAAGCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3738
|
|
|
Db 3421 CTACCCAAAGCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3480
|
|
|
QY 3739 GCCCAGAACGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3798
|
|
|
Db 3481 GCCCAGAACGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3540
|
|
|
QY 3799 GAGGAGCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3858
|
|
|
Db 3541 GAGGAGCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3600
|
|
|
QY 3859 TATGTAAGCTGAGACAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3918
|
|
|
Db 3601 TATGTAAGCTGAGACAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3660
|
|
|
QY 3919 CCCCCTCAGGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3978
|
|
|
Db 3661 CCCCCTCAGGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3720
|
|
|
QY 3979 TGA 3981
|
|
|

Db 3721 TGA 3723

RESULT 4
AA239025
ID AA239025 standard; cDNA, 5144 BP.
XX
XX AA239025;
AC
XX
XX 28-FEB-2000 (first entry)
DT
XX
XX Mouse Esell coding sequence.
DE
XX
XX Mouse; murine; Esel1; Esel2; endocytosis; vesicular trafficking;
KW regulation; actin cytoskeleton; detection; cancer; infection;
KW EH-domain and SH3-domain regulator of endocytosis; anticancer;
KW antiproliferative; antiviral; ss.
XX
XX Mus sp.
OS
XX WO955728-A2.
XX
XX 04-NOV-1999.
XX
XX 27-APR-1999; 99WO-CA00375.
XX
XX 27-APR-1998; 98CA-2230201.
XX 05-FEB-1999; 99US-0118739.
XX
XX (HSCR-) HSC RES & DEV LP.
XX
XX Egan SE, Wang W, Sengar A;
PI
XX MPI: 2000-052802/04.
XX P-PDB; AA157449.
XX
XX New nucleic acid encoding Esel1 and 2 proteins, involved in regulation
PT of endocytosis, used e.g. for treating cancer or preventing viral
PT infection -
XX
XX Claim 6; Page 59-62; 99pp; English.

The present invention specifically describes mammalian Esel1 and 2 proteins (I) and their splice variants (Ese = EH-domain and SH3-domain regulator of endocytosis). (I) are involved in regulation of clathrin-mediated endocytosis (as a complex with Eps15 protein), vesicular trafficking and actin cytoskeleton. Generally (I) (or its (ant)agonists, mimetics, fragments and inactive mutants); (I)-specific antibodies (Ab); sequences antisense to the (I) polynucleotide; agents that downregulate expression of Ese genes or antagonists of an Ese binding partner are used to treat diseases associated with undesirable endocytosis and resulting changes in cellular function. Particularly overexpression of Esel1 is used to block clathrin-mediated endocytosis in vivo or in cell cultures, while administration of (I) is used to promote endocytosis of selected cells. (Ant)agonists of (I) or Ab are used to suppress abnormal proliferation of cells that can be stimulated to proliferate by a growth factor receptor; and similar compounds (also inactive Ese mutants) can be used to prevent viral infection. Endocytosis may also be regulated, in vivo or in cell cultures, by forming an Ese-Eps15 complex, then binding dynamin to the complex. Generally conditions that can be treated include cancer; abnormal cell division or migration; viral infection; or abnormal receptor signaling, tissue development or synaptic transmission. The present sequence represents mouse Esell coding sequence.

Sequence 5144 BP; 1447 A; 1298 C; 1415 G; 984 T; 0 other;

Query Match 71.6%; Score 3639.4; DB 21; Length 5144;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3640; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 259 ATGGCTCAAGTTCCACACCTTTGGTGTAGCTGTGATGTCTGGCCATACTGTGAG 318
DB 1 ATGGCTCAAGTTCCACACCTTTGGTGTAGCTGTGATGTCTGGCCATACTGTGAG 60

QY 319 GAAAGGGCCAGCATGATGACAGAGTTCCCTTAGCTGAAGCCGATGACGGGATTTATTACT 378
DB 61 GAAAGGGCCAGCATGATGACAGAGTTCCCTTAGCTGAAGCCGATGACGGGATTTATTACT 120
QY 379 GGTGATCAAGCAGGAACTTTTTCATCATCTGGGTTACTCAAGCTGTCTTACACAA 438
DB 121 GGTGATCAAGCAGGAACTTTTTCATCATCTGGGTTACTCAAGCTGTCTTACACAA 180
QY 439 ATATGGGGCTTAGCGGACATGATTAAGATGGAAGATGATCAAGTGAATTTTCATA 498
DB 181 ATATGGGGCTTAGCGGACATGATTAAGATGGAAGATGATCAAGTGAATTTTCATA 240
QY 499 GCCATGAAGCTATCAACTGAAGTAAAGGATATCAAGTCCCTCCATCACTTCCCTCC 558
DB 241 GCCATGAAGCTATCAACTGAAGTAAAGGATATCAAGTCCCTCCATCACTTCCCTCC 300
QY 559 GTCATGAACACAGCAACAGTGGCTATTTCCAGTGACACAGCATTTGTATAGAGGANT 618
DB 301 GTCATGAACACAGCAACAGTGGCTATTTCCAGTGACACAGCATTTGTATAGAGGANT 360
QY 619 GCTAGCATGCCACCACTCAAGCTGTGCTCTGTGCCAATGGGCTCCATTCAGTTGTT 678
DB 361 GCTAGCATGCCACCACTCAAGCTGTGCTCTGTGCCAATGGGCTCCATTCAGTTGTT 420
QY 679 GGAATGTCACCCCTTAGTATCTTCTGTCCTCCAGAGAGAGTGGCTCCCTGGCTAAC 738
DB 421 GGAATGTCACCCCTTAGTATCTTCTGTCCTCCAGAGAGAGTGGCTCCCTGGCTAAC 480
QY 739 GGGGCTCTCCCGTATACAGCTGTGCTGGCTGGTGGCATCTGACGCCACATGGCCA 798
DB 481 GGGGCTCTCCCGTATACAGCTGTGCTGGCTGGTGGCATCTGACGCCACATGGCCA 540
QY 799 AAGAGTTCTTCTTCAAGCATCTGCTCAGGCTCACAATTAAACATTAAGTTAACAAG 858
DB 541 AAGAGTTCTTCTTCAAGCATCTGCTCAGGCTCACAATTAAACATTAAGTTAACAAG 600
QY 859 GCACATCATTCGATGTGGCCAGGCGCCCTCCAGAGAGAGATGGGCTGTGCTACATCA 918
DB 601 GCACATCATTCGATGTGGCCAGGCGCCCTCCAGAGAGAGATGGGCTGTGCTACATCA 660
QY 919 TCAAGCTGAATATACAGGACAGTTATTTCAACAGCCACGACAAACTATGATGACACTTA 978
DB 661 TCAAGCTGAATATACAGGACAGTTATTTCAACAGCCACGACAAACTATGATGACACTTA 720
QY 979 ACAGGTCCTCCAGCAAGAACTATTTCTCATGCAATCAAGTTACCCAGGCTCAGTGGCT 1038
DB 721 ACAGGTCCTCCAGCAAGAACTATTTCTCATGCAATCAAGTTACCCAGGCTCAGTGGCT 780
QY 1039 TCAATATGAATCTTTTGAACATTCATGATGAAGTAAACTCATGCAAGAAATTTATC 1098
DB 781 TCAATATGAATCTTTTGAACATTCATGATGAAGTAAACTCATGCAAGAAATTTATC 840
QY 1099 CTAGCTATGACCTTATGATGTGTCCATGTCTGTGTCAGCCACTGCGCCGCTCTGCT 1158
DB 841 CTAGCTATGACCTTATGATGTGTCCATGTCTGTGTCAGCCACTGCGCCGCTCTGCT 900
QY 1159 CCAGATATCATCTCTCTTCTTGAAGAGTTGCTCCGAGAGTGGATGTCCGTCAATA 1218
DB 901 CCAGATATCATCTCTCTTCTTGAAGAGTTGCTCCGAGAGTGGATGTCCGTCAATA 960
QY 1219 AGCTTTCTTCTGTGTGATCAAGAGCTGCTGAGAGCCGTGCTCAGAGATGAGACAG 1278
DB 961 AGCTTTCTTCTGTGTGATCAAGAGCTGCTGAGAGCCGTGCTCAGAGATGAGACAG 1020
QY 1279 CCAGAGAGAAACTGCTGTGATTCATTTGAAGTAAAGACGGGAACTTCAGAGGAGGC 1338
DB 1021 CCAGAGAGAAACTGCTGTGATTCATTTGAAGTAAAGACGGGAACTTCAGAGGAGGC 1080
QY 1339 AGTGTGAGCTGGAGAGAGCCGCAAGCGCTCTTGAAGCAGACGCAAGAGCAGAG 1398
DB 1081 AGTGTGAGCTGGAGAGAGCCGCAAGCGCTCTTGAAGCAGACGCAAGAGCAGAG 1140

Oy	1339	CGGTTGGCTCAGCTGGAGCGCGCCGACGACGAGAGAGAAAGACGGGACCGCCAGAGACG	1458
Db	1141	CGGTTGGCTCAGCTGGAGCGCGCCGACGACGAGAGAGAAAGACGGGACCGCCAGAGACG	1200
Oy	1459	GAGCCAAAGCGGACAGCTGAGAGAAAGCAGCTGGAGAAAGCAGCGGAGCTGGAGCGG	1518
Db	1201	GAGCCAAAGCGGACAGCTGAGAGAAAGCAGCTGGAGAAAGCAGCGGAGCTGGAGCGG	1266
Oy	1519	CAGCGAGAGAGAGAGAGAGAGATCGAGAGCGCGAGGCGCAAAACGGGAACTG	1578
Db	1261	CAGCGAGAGAGAGAGAGAGAGATCGAGAGCGCGAGGCGCGCAAAACGGGAACTG	1320
Oy	1579	GAAAGGAGCCACCACTTGAATGGGAAACGGAAACCGGACACAGAACTCTGAAATGAGG	1638
Db	1321	GAAAGGAGCCACCACTTGAATGGGAAACGGAAACCGGAAACCGGAACTCTGAAATGAGG	1380
Oy	1639	AACAAGGAGCAGGAGGAGCGCGTGCTCGAAAGGCAAGAGGAACTCTGAGATTGAG	1698
Db	1381	AACAAGGAGCAGGAGGAGCGCGTGCTCGAAAGGCAAGAGGAACTCTGAGATTGAG	1440
Oy	1699	TTAGAAGCTCTGAATGACAAAAAGCATCAGCTGAAGGAAACTTCAGGATATCAGTGT	1758
Db	1441	TTAGAAGCTCTGAATGACAAAAAGCATCAGCTGAAGGAAACTTCAGGATATCAGTGT	1500
Oy	1759	CGACTGGCAACCCAGAGGCAAGAAATTGAGAGCAGCAACAGCTTAGAGACTTAAGATT	1818
Db	1501	CGACTGGCAACCCAGAGGCAAGAAATTGAGAGCAGCAACAGCTTAGAGACTTAAGATT	1560
Oy	1819	GCTGAATATCACCACCTTACAGAGAGATTGGACGAATCTCAGCAATGCTTGGAAACTT	1878
Db	1561	GCTGAATATCACCACCTTACAGAGAGATTGGACGAATCTCAGCAATGCTTGGAAACTT	1622
Oy	1879	ATTCAGAGAAACAGATATCTCAGTGAACGATTAAAAACAAGTCAGCAGAACAGTTTGAT	1938
Db	1621	ATTCAGAGAAACAGATATCTCAGTGAACGATTAAAAACAAGTCAGCAGAACAGTTTGAT	1680
Oy	1939	AGAGACTGGCTTTACCTCAAAAGAGCTTTGAAAGCAAGAGAGCTGGCCGGCAGAG	1998
Db	1681	AGAGACTGGCTTTACCTCAAAAGAGCTTTGAAAGCAAGAGAGCTGGCCGGCAGAG	1740
Oy	1999	CTCCGGGAGCAGCTGAGCGAGGTGGAGAGAGACAGGTCAACCTCAGAGAGATTGAT	2058
Db	1741	CTCCGGGAGCAGCTGAGCGAGGTGGAGAGAGACAGGTCAACCTCAGAGAGATTGAT	1800
Oy	2059	GTTTTCAAACAACAGCTGAGAGAACTGAGAGAGATACATAGCAAAACAGAACTCCAGAG	2118
Db	1801	GTTTTCAAACAACAGCTGAGAGAACTGAGAGAGATACATAGCAAAACAGAACTCCAGAG	1860
Oy	2119	CAGAGGTCCCTGAGGCGAGCGGACTGAAGCAGAAAGCAGAGAGAGAGAGCTGGAG	2178
Db	1861	CAGAGGTCCCTGAGGCGAGCGGACTGAAGCAGAAAGCAGAGAGAGAGAGCTGGAG	1920
Oy	2179	TTAGAGAAAGCAAAAGAAAGCCCTCAGAGACGAGTTCAAGAAAGGGAACAAGATGGCTG	2238
Db	1921	TTAGAGAAAGCAAAAGAAAGCCCTCAGAGACGAGTTCAAGAAAGGGAACAAGATGGCTG	1980
Oy	2239	GAGCATGTGCACAGAGAGAGCAGCCGCCCCCGGAAACCCCAAGAGAGACAGACTG	2298
Db	1981	GAGCATGTGCACAGAGAGAGCAGCCGCCCCCGGAAACCCCAAGAGAGACAGACTG	2040
Oy	2299	AAGAGGGAAGACAGTGTCAAGAAAGAGAGGCGGGAAGAGAGCCAAAGCCGGAATTGCA	2358
Db	2041	AAGAGGGAAGACAGTGTCAAGAAAGAGAGGCGGGAAGAGAGCCAAAGCCGGAATTGCA	2100
Oy	2359	GACAAAGCAGAGTGGCTTTTCCATCCGATCAGAGACCAGCTAAGCTGGCCACCAAGCA	2418
Db	2101	GACAAAGCAGAGTGGCTTTTCCATCCGATCAGAGACCAGCTAAGCTGGCCACCAAGCA	2160
Oy	2419	CCCTGGTCTACACAGAGAAAGCCGCTTACATTTCTGCACAGAGAGGTGTAAGTG	2478
Db	2161	CCCTGGTCTACACAGAGAAAGCCGCTTACATTTCTGCACAGAGAGGTGTAAGTG	2222
Oy	2479	GTAATTTACGAGCGCTGATCCCTTTGAAATCAGAGAGTCAGATGAGATCCATTCAG	2538

Db	2221	GTATATTACGAGGCTGTACCTCTTGTAATCCAAAGTCAAGATGAGATACATCCAG	2280
Qy	2539	CCAGAGATATATGTCATGTGTGATGAAAGCCAGACTGAGAGCCAGAGATGGCTTGGAGA	2598
Db	2281	CCAGAGATATATGTCATGTGTGATGAAAGCCAGACTGAGAGCCAGAGATGGCTTGGAGA	2340
Qy	2599	GAGCTGAAGGGAGAGACGGGATGGTTCCTCTGCAACTATGTGCAAAAAGATTCCGAAAT	2658
Db	2341	GAGCTGAAGGGAGAGAGCGGATGGTTCCTCTGCAACTATGTGCAAAAAGATTCCGAAAT	2400
Qy	2659	GAGGTCCCATCTCAGGCCAAACGAGTGACCGATGTGCATCTGGCCCTGCCCAACTG	2718
Db	2401	GAGGTCCCATCTCAGGCCAAACGAGTGACCGATGTGCATCTGGCCCTGCCCAACTG	2460
Qy	2719	GCTCTGGGTGAGACCCCTGCTCTTTTCCAGATGACCTTTCTGAGCCCTCCACAACCCC	2778
Db	2461	GCTCTGGGTGAGACCCCTGCTCTTTTCCAGATGACCTTTCTGAGCCCTCCACAACCCC	2520
Qy	2779	AACAACTGGGCGAGCTTCAGTTCCAGTCGCGCCAGCAAGCTCAACGAGAACCGAAGACG	2838
Db	2521	AACAACTGGGCGAGCTTCAGTTCCAGTCGCGCCAGCAAGCTCAACGAGAACCGAAGACG	2580
Qy	2839	GACAACTGGGATACGTGGCGGCTCAGCCTTCTGTACCCGATCCTAGTGGCGGCAAGTA	2898
Db	2581	GACAACTGGGATACGTGGCGGCTCAGCCTTCTGTACCCGATCCTAGTGGCGGCAAGTA	2640
Qy	2899	CGGAGAGATAGCCTTTACCCCGACAGCAGCAGCAGCTGGCTCTCCCATCTCCCGTCTG	2958
Db	2641	CGGAGAGATAGCCTTTACCCCGACAGCAGCAGCAGCTGGCTCTCCCATCTCCCGTCTG	2700
Qy	2959	GGCCAGGTTGAAAAGTGGAAAGGCTCAACAGCGCAAGCCCTGTATCCCTGGAGAGCCAA	3018
Db	2701	GGCCAGGTTGAAAAGTGGAAAGGCTCAACAGCGCAAGCCCTGTATCCCTGGAGAGCCAA	2760
Qy	3019	AAAGCAACCACTTAAATTTTAAACAAAAGTACGTCATCACCGTTCTGGAAACACCAAGAC	3078
Db	2761	AAAGCAACCACTTAAATTTTAAACAAAAGTACGTCATCACCGTTCTGGAAACACCAAGAC	2820
Qy	3079	ATGTGTGTTTGGAGAAAGTCAAGGTTCAGAAAGGTTGGTCCCAAGTCTTACGTGAAA	3138
Db	2821	ATGTGTGTTTGGAGAAAGTCAAGGTTCAGAAAGGTTGGTCCCAAGTCTTACGTGAAA	2880
Qy	3139	CTCATTTCCAGGGCCGTAAAGGAAATCCACAAGCATCGATCTGGCCCTACTGAAAGTCT	3198
Db	2881	CTCATTTCCAGGGCCGTAAAGGAAATCCACAAGCATCGATCTGGCCCTACTGAAAGTCT	2940
Qy	3199	GCTAGTCTTAAAGAGATGGCTTCCCCCGGCCCAAGCCAGCATTTCCCGAGAAAGATT	3258
Db	2941	GCTAGTCTTAAAGAGATGGCTTCCCCCGGCCCAAGCCAGCATTTCCCGAGAAAGATT	3000
Qy	3259	ATTGCAATGTACATACGAGATCTGAGACAGAGATTTAACTTTACGTTACAGCAAGGGAT	3318
Db	3001	ATTGCAATGTACATACGAGATCTGAGACAGAGATTTAACTTTACGTTACAGCAAGGGAT	3060
Qy	3319	GTGATTTGTATTACCAAGAAAGATGTGTGACTGTGTGACGGGACGGTGGCGACAAAGTCC	3378
Db	3061	GTGATTTGTATTACCAAGAAAGATGTGTGACTGTGTGACGGGACGGTGGCGACAAAGTCC	3120
Qy	3379	GGAATCTTCCCTTCTAACTATGTAGGCTTAAAGATTCAGAGGGCTCTGGAACGTCTGGG	3438
Db	3121	GGAATCTTCCCTTCTAACTATGTAGGCTTAAAGATTCAGAGGGCTCTGGAACGTCTGGG	3180
Qy	3439	AAAAACAGGAGATTAGGAAAAAAAACCTGAATTGCCAGGTTATGTCTCTCAAGCTGCT	3498
Db	3181	AAAAACAGGAGATTAGGAAAAAAAACCTGAATTGCCAGGTTATGTCTCTCAAGCTGCT	3240
Qy	3499	ACTGGTCCCGAACAACCTACCTGAGCTCTGAGGAGAGCTGATTCTGATCCGAAAAAAGAC	3558
Db	3241	ACTGGTCCCGAACAACCTACCTGAGCTCTGAGGAGAGCTGATTCTGATCCGAAAAAAGAC	3300
Qy	3559	CCAGGTGATGTGTGGAGAGAACTGCAAGCTCGAGGAGAAAAAGCCGACATAGCGTGG	3618

Db 3301 CCAAGTGTAGTGGGAGAGAGAACTGCAAGCTCGAGGAGAAAAAGCCGACATAGGGTGG 3360
 QY 3619 TTTCAGCAAAATATATGCTAACTTTAAGCCCGGAAAGCAAGAAATACCCCAACTGAG 3678
 Db 3361 TTTCAGCAAAATATATGCTAACTTTAAGCCCGGAAAGCAAGAAATACCCCAACTGAG 3420
 QY 3679 CTACCCAGAACCCGACAGTGCAGCCAGAGTGTGCAGAGTATCGGATGTACATTAACACC 3738
 Db 3421 CTACCCAGAACCCGACAGTGCAGCCAGAGTGTGCAGAGTATCGGATGTACATTAACACC 3480
 QY 3739 GCCCGAAGACATGACCACTAGCTTACAGAAAGGCCAATCATCACTCTCAACAG 3798
 Db 3481 GCCCGAAGACATGACCACTAGCTTACAGAAAGGCCAATCATCACTCTCAACAG 3540
 QY 3799 GAGGACCCGACAGTGTGTGAAAGAGAAAGTCAAGTGGGCAAGTTGGGCTTTCCATCCAT 3858
 Db 3541 GAGGACCCGACAGTGTGTGAAAGAGAAAGTCAAGTGGGCAAGTTGGGCTTTCCATCCAT 3600
 QY 3859 TATGTAAAGCTGACCAAGACATGAGACCCGACAGCAATG 3899
 Db 3601 TATGTAAAGCTGACCAAGACATGAGACCCGACAGCAATG 3641

RESULT 5
 AA234571
 ID AA234571 standard; cDNA; 5458 BP.
 AC AA234571;
 DT 01-FEB-2000 (first entry)
 DE Human SH3D1A cDNA clone 21.
 XX
 XX SH3D1A gene; human; Down's syndrome; leukaemia; cancer;
 KW megakaryocytic abnormality; myeloproliferative disorder;
 KW platelet disorder; neural disorder; thrombocytopenia;
 KW haematopoietic disorder; cognitive dysfunction; microcephaly;
 KW lissencephaly; colpocephaly; holoprosencephaly; diagnosis; therapy;
 ss.
 OS Homo sapiens.
 XX
 XX
 FH Key Location/Qualifiers
 FT CDS 267..3929
 FT /*tag= a
 XX
 XX WO9953062-A2.
 PD 21-OCT-1999.
 PD
 PF 16-APR-1999; 99WO-US08371.
 XX
 XX 16-APR-1998; 98US-0082007.
 XX
 XX (CEDA-) CEDARS SINAI HEALTH SYSTEM.
 XX
 XX Korenberg JR, Chen X;
 DR WPI; 1999-633829/54.
 DR P-PSDB; AAY32155.
 XX
 XX Nucleic acid from the human SH3D1A gene and its products, useful for
 PT the diagnosis and treatment of myeloproliferative disorders and
 PT leukaemia
 XX
 PS Claim 2; Fig 8; 99p; English.
 CC This is the nucleotide sequence of full-length cDNA (clone 21)
 CC corresponding to a novel human SH3 gene, termed the SH3D1A gene,
 CC that contributes to the development of platelets and the
 CC pathogenesis of leukaemias, both in general and in particular those
 CC involving the megakaryocytic lineage. The SH3D1A gene maps to the
 CC small candidate region for low platelets on chromosome 21.

CC Sequencing of 5 different sizes of cDNA clone from foetal brain
 CC (see AA234570-74) suggests that at least 3 isoforms exist. The
 CC invention provides methods for the diagnosis and treatment of
 CC megakaryocytic abnormality, myeloproliferative disorder, platelet
 CC disorder, acute leukaemia, neural disorders, thrombocytopenia,
 CC platelet disorder on chromosome 21, low platelets in deletion for
 CC 21, association of gains in chromosome 21 with leukaemias, neural
 CC abnormalities, dysfunctions and disorders including brain
 CC malformations and corresponding cognitive dysfunction,
 CC microcephaly, lissencephaly, and colpocephaly. Methods are also
 CC provided for: suppressing cells unable to regulate themselves;
 CC screening for a somatic alteration in the SH3D1A gene; monitoring
 CC the progress and adequacy of a treatment; monitoring tumour risk
 CC progress or megakaryocytic abnormality, myeloproliferative disorder,
 CC haematopoietic disorder, platelet disorder or leukaemia; and
 CC treatment of a subject (including a prenatal subject) having
 CC megakaryocytic abnormality, myeloproliferative disorder,
 CC platelet disorder, leukaemia or neural disorder using a
 CC nucleic acid that expresses SH3D1A or its antisense nucleic acid.
 XX

SQ Sequence 5458 BP; 1671 A; 1201 C; 1307 G; 1279 T; 0 other:
 Query Match 69.4%; Score 3529.4; DB 20; Length 5458;
 Best Local Similarity 83.7%; Pred. No. 0;
 Matches 4278; Conservative 0; Mismatches 766; Indels 69; Gaps 22;

QY 7 GAGGAGAGTGAAGCGGCGCGGAGGCGCGAGCTTGCTCTCGTGTACCGCGGCT 66
 Db 25 GAGGAGAGTGAAGCGGCGCGGAGGCGCGAGCTTGCTCTCGTGTACCGCGGCT 84
 QY 67 GCGAAGGAGCATCCCGAGCGGCGCTCCGGGAGCGCGGAGCGGAGCGGCGGCG 126
 Db 85 CCGAGGAGAAATCCCGAGCGGCGCTCCGGAGC-----GACAGAGAGCGGCGG 133
 QY 127 GGGATGTGTGCGCGCGCTCGGACTCGCGTCTCGC-GCGGCGTGCAGCTGCACTGA 185
 Db 134 GGGATGTGTGCGCGGCGCTCGCGTCTCGCGTCTCGCAGCGGCGGAGGCGCACTGA 193
 QY 186 TTTGTGTAGGGGCGGCGCGCGCGACCGCGCGAGATGAGCGCTGATCAGCAAGTGA 245
 Db 194 TTTGTCTCCGCGGCGCGAGCGCGGACCGCGCGAGATGAGCGCTGATCAGCAAGTGA 253
 QY 246 ACGTAATGAAACATAGGCGCTCACTTCCCAACCTTCGGTGTACCGAGATGTCCGGCG 305
 Db 254 AAGTAACAGAACATAGGCGCTCACTTCCCAACCTTTCGGTGTACCGAGATGTCCGGCG 313
 QY 306 CATTAAGTGAAGAAAGGCGGAGCATGACGAGAGTTCCTTAGCTGAAGCCGATAGC 365
 Db 314 CATTAAGTGAAGAAAGGCGGAGCATGACGAGAGTTCCTTAGCTGAAGCCGATATC 373
 QY 366 GGGATTTTATCTGTGTATCAAGCGAGAACTTTTTCATCTGGGTTACTCAGCC 425
 Db 374 TGGATTTCACTGTGTATCAAGCTAGAAACTTTTTCATCTGGGTTACTCAGCC 433
 QY 426 TGTCTTAGACAAATATGGGCGCTAGCGGACATGAATACATGAGATGAGATCAAGT 485
 Db 434 TGTCTTAGACAAATATGGGCGCTAGCGGACATGAATATGATGAGATGAGATCAAGT 493
 QY 486 GGAATTTTCATAGCCATGAAGCTTATCAAACTGAAGCTACAGAGATACAGCTCCCTC 545
 Db 494 GGAATTTTCATAGCCATGAAGCTTATCAAACTGAAGCTACAGAGATACAGCTCCCTC 553
 QY 546 CACACTTCCCTGTGCATGAAGAAAGAGAACAGAGTGTATTTTCAGTGCACGCAATTGG 605
 Db 554 TGCACCTTCCCTGTGCATGAAGAAAGAGAACAGAGTGTATTTTCAGTGCACGCAATTGG 613
 QY 606 TATAGAGGAGATTGTAGATGCCACGCTACAGCTGTGCTCTGTGCCAATGGGCTC 665
 Db 614 TATAGAGGATATGCCAGATGCCACCGCTTACAGCTGTGCTCTGTGCCAATGGGATC 673
 QY 666 CATTCAGTGTGTGAATGTCTCCACCTTAGATCTTCTGTCCCTCCAGCAGAGTGC 725
 Db 674 CATTCAGTGTGTGAATGTCTCCACCTTAGATCTTCTGTCCCAAGCAGAGTGTGC 733

OY	726	TCCTTGECTAACGGGGGCTCTCCCGTCAATACGCTCTGCTTGCTTGCGCATCTCGC	785
Db	734	CCCCCTGGCTAAACGGGGGCTCCCGCTGTATACAACTCTGCTGATTTGGCTCACTCGC	793
OY	786	AGCCACATGGCCAAAGAGTTCTTCTTGACAGATCTGTGCAGGGGTCACAATTAAAC	845
Db	794	AGCCACATGGCCAAAGAGTTCTTCTTATAGATCTGTGCAGGGGTCACACTTAAC	853
OY	846	TAAATTACAAAGGCAAGCAATCATCTGATGTGGCCAGCGCCCTCCAGACAGAAATGGGC	905
Db	854	TAAATTAACAAAGGCAAGCAATTTGAATGTGGCCAGTCTCCACAGTGGGAGATGGGC	913
OY	966	TGTGCTCAGTCATCAAGGCTGAAATACAGGAGATTATCAACAGCCACGACAAACTAT	965
Db	914	TGTTCTCAGTCATCAAGCTGAAATACAGGCAATTAATTCATATGTCATGACAAACTAT	973
OY	966	GAGTGGACATTTAACAGGTCCCGACGCAAGAACTATCTCATGCAATCAAGTTTACCCA	1025E
Db	974	GAGTGGACATTTAACAGGTCCCGACGCAAGAACTATCTTATGCAAGTTCATGCA	1033
OY	1026	GGCTCAGCTGGCTTCAATATGGAAATCTTTCTGACATTTGATCAAGATGGAAAACTCATG	1085B
Db	1034	GGCTCAGCTGGCTTCAATATGGAAATCTTTCTGACATTTGATCAAGATGGAAAACTTACAG	1093B
OY	1086	AGAAATTTTATCCTAGCTATGCACTTAATGATTTGATTTGCCATGTGTGCTACGCACTGCC	1145A
Db	1094	AGAGAAATTTATCCTAGCTATGCACTTATGATTTGATTTGCTGACATGTCGCAACCTGCC	1155A
OY	1146	GCCGCTCTGCTCTCCAGAAATACCTCCCTCTTCTTCAAGAAATTTGGCTCCGCGCATGG	1205B
Db	1154	ACCTGCTCTGCTCTCCAGAAATACCTCCCTCTTCTTCAAGAAATTTGATCTGGCGATGG	1213B
OY	1206	GATGTCGCTCAATACCTCTCTTCTGTGTGATCAGAGGCTGCGTGAAGCCGTGCTCAGA	1265B
Db	1214	TATATCTGCTCAATACCTCAACATCTGTGATATAGAGGCTTACCAAGAGAACTCAAGTTTGA	1273B
OY	1266	GGATGAGCAGCAGC---CAGAGAAAGAACTGCTGTGACATTTGAAATTAAGAACGGGA	1322B
Db	1274	AGATGAACAACAATTAAGAAAGAAATTAACCTGTAACTGTTGAAGATTAAGAACGGGA	1333B
OY	1323	GAACTTGAGGAGGCAAGTGTGAACTGTGAAAGGCGCGCAAGGCTCTTGGACACAGA	1382B
Db	1334	GAACTTTGAACCTGTGCAACTGTGAATGTGAGAAAGAAAGCAAGCTCTCTGAAACAGA	1393B
OY	1383	GCGCAAGACAGAGAGCGGTTGGCTCAGCTGGAGCGCGCCGACGACAGAGAGAGAAAGACG	1442B
Db	1394	GCGCAAGACAGAGAGCGGCTGGCCAGCTGAGAGCGGCGGAGCGAGAGAGAAAGACG	1455B
OY	1443	GGAAGCGCAAGACAGAGAGCCAGAGCGGCACTTGAAGCTTGAAGAGCAAGCTGTGAGAGCA	1502B
Db	1454	TGAGGCGCAAGAGAGAGCGCAAGAGCACTGTGAACCTTGAAGAGCAACTGTGAAACCA	1513B
OY	1503	GCGGAGCTGGAGCGGACAGAGAGAGAGAGAGAGAGAGATTCGAGAGGCGGAGAGC	1567B
Db	1514	GCGGAGCTTGAACCGGACAGAGAGAGAGAGAGAGAGAGAGAAATTTGAGAGGCGGAGAGC	1573B
OY	1563	CGCAAAACGGGAATCTGAAAGGCGACGCAACTTGAATGTGAACGGAACTCGAGACAGGA	1622B
Db	1574	TGCAAAACGGGAATCTTGAAGGCAAGCAACTTGAATGTGAGAAAGGAACTCGAAGGCAAGA	1633B
OY	1623	ACTCCTGAATCAAGAGAACAGAGACAGAGAGGCAACCTGTGCTTGAAGGCAAGAGAGAA	1682B
Db	1634	ACTACTAATCAAAAGAACAAAGAACAGAGACATGTGTGACTGAAAGCAAAAGAAA	1693B
OY	1683	GACTCTGAGTTTGATTAAGAGCTCTGATATGACAAAGAAAGCACTCAGCTAGAGAGAAACT	1744B
Db	1694	GACTTTGGAAATTTGAATTAAGAGCTCTAATATGATTAAGAAAGCACTCAACTAGAGAGAACT	1755B
OY	1743	TCAGAGATTCAGGTGTCCATGTGCAACCCAGAGCGCAAGAAATTTGAGAGCAAGACAAATC	1802B
Db	1754	TCMAATATTCAGATTCGATTCGATTCGACCAACCAAGGCAAGAAATTTGAGACACAAACAAATC	1813B

QY	1803	TAGGAGCTAAGATTGCTGAATTCACCCACTTACACAGACAGTTGAGGAATTCAGCA	1862
Db	1814	TAGGAGTTGGAATTCGCCGAATCACCCACTTACACGAACAATTTACGAAATTCAGCA	1873
QY	1863	AATGCTTGGAAAGCTTATTTACAGAGAAACAGATACTACAGTGAACAGTTAAACAAATCCA	1932
Db	1874	AATCTTGGAAAGCTTATTTCCAGAAAAACAGATATCTAAATGACCAATTTAAACAAAGTTCA	1933
QY	1923	GCAGAACGTTTGCAATGAGACTGCTTCTTACCTCCMAAAGACCTTGGAAAGCAAGAA	1982
Db	1934	GCAGAACGTTTGCAAGAGATTCATCTTTGTAACCTTTAAAGAGCCTTAGAAGCAAAAGA	1993
QY	1983	GCTGGCCCCGAGACGCTCCGGGAGCAGCTGGAACGAGAGTGGAGAGAGACACAGTCAAA	2042
Db	1994	ACTGAGCTCGGAGACCTTACGAGCCAACTGGAGTAAGTGGAGAAAGAAATCTAATATCAA	2053
QY	2043	GCTCAGAGAGTTGATGTTTTCAACAACAGCTGAAGAACTGAGAGAGATATCATACAA	2102
Db	2054	ACTCAGAGAGTTATATTTTCAATTAATCAGCTGAAGAACTTAAGAGAAATACAAATAA	2113
QY	2103	ACAGCAATCCAGAAAGCAGAGGTCCCTGAGAGCACCGCACTGAAAGCAAGAAAGACAGA	2162
Db	2114	GCAACAATCCAGAAAGCAAAAGTCCATGAGAGGCTGAAGACTGAACAGAAAGAAACAAGA	2173
QY	2163	GAGGAAAGCTCGAGATTAGAGAAACAAAGAAAGACGCTCAGAGACGAGTTCCAGAAAG	2222
Db	2174	ACGAAAGATCTAATGAATTAGAAAAACAAAAAGAAAGAGCCCAAGAGAGAGCTCGAAGAG	2233
QY	2223	GGACAGCAATAGGTGGAGCATGAGCAGCAGGAG- -GAGCAGCAGCGCCCGCGGAATCC	2279
Db	2234	GGACAGAGTGGCTGGAGCATGTCAGCAGAGAGACGAGCATAGAGATCAGAAACT	2293
QY	2280	CCACGAGAGAGACGAGCTGAAGAGGAGAAAGCAGTGTCAAGAAAGAAAGAGCGGAAGAG	2339
Db	2294	CCACGAGAGAGAAATCTGAAGAAAGGAGAGAGTGTCAAAAGAAAGATGGCGAGAAAA	2353
QY	2340	AGCCAAAGCCGAAAAATGCAAGACAGCAGAGTCCGCTTTTCATCCGATCCAGACCAAC	2399
Db	2354	AGGCAAAACAGAAACCAAGACAGCAGTGGCTTTTCCATCAACCAAGAAACAGC	2413
QY	2400	TAACTGGCCACCAGGACCCCTGGTCTACACAGAGAAAGCCCGCTTACATTCCTGAC	2459
Db	2414	TAAAGCAGCTGTCCAGGACCCCTGTCTCATGCAAGAAAGATCTCATTAATTCATTTCTGC	2473
QY	2460	ACAGAGAGTGTAAAGTGTATTTAACGAGCGCTGTACCCCTTTGAATCCAGAAATCTCA	2519
Db	2474	ACAGAGAAATTTAAAGTGTATTTACCGGACCTGTACCCCTTTGAATCCAGAAACGA	2533
QY	2520	CGATGAGATCCATCCAGCCAGGAGATATATCAT-----GGTGATGA	2564
Db	2534	TGATGAATCATCTATCCAGCCAGGAGACATATGATGTTAAAGGGAATGGGTGATGA	2593
QY	2555	AAGCCAGCTGGAAAGCCAGATGGCTTTGAGAGAGACTGAABAGGAAAGCCGGAATGGTT	2624
Db	2594	AAGCCAACTGGAAACCCCGCTGGCTTGGAGGAAATTAAGAGAAAGAACAGGGGTGGTT	2633
QY	2625	CCCTGCAAACTATGTCAGAAAGATTCAGAAAGATGAGGTTCCTCACTCCAGCCAAACAGT	2684
Db	2654	CCCTGCAAACTATGACAGAAATATCCAGAAAGATGAGGTTCCTCACTCAAGTAAACAGT	2713
QY	2685	GACCGATCTGACATCTGCCCTTGCCTCCCAACTGGCTCTGCGTAGACCCCTGCTCTTT	2744
Db	2714	GACGATTTCAACATCTGCCCTTGCCTCCCAACTGGCTTGGGTGAAGACCCCGCTCTTT	2773
QY	2745	GCAGTGAATCTTTCTGAGCCCTTCAACAACCCCAACAACCTGGGCAAGCTTCAATTCAC	2804
Db	2774	GGCAGTAACTCTTTCAAGCCCTTCAAGACCCCTTAATTAACCTGGGCGAATTCACCTCCAC	2833
QY	2805	GTTGGCCCAAGCAGCTCAACAGGAAGCCAGAAACGGAACAATGGGATTCAGTGGGGGCTCA	2864
Db	2834	GTTGGCCCAAGCAGATGAAGAAACAGAAACGGAATACCTGGGATGATGGGCAAGCCCA	2893
QY	2865	GCTTCTCTGACCGTAACTATGCTGGCAGATTACCGCAGAGATCAGCTTTTACCCAGC	2924

Db 2894 GGCCTCTCAACGGTTCAGAGTCCGCGCAGTTAAAGCAGAGGCTCCGCTTTACTCCAG 2953
QY 2925 CACAGCACTGGCTCCTCCCATCTCCCGCTCGGCGCAGGGTGAAGAGTGAAGGCT 2984
Db 2954 CAGGCGACCTGGCTCCTCCCGCTCCTGTGTAGGCGAGGGTGAAGAGTGAAGGCT 3013
QY 2985 ACNAGCGAAGCCCTGTATCCCTGAGAGCGCAAAAAAGAACCACTTAATTTTAA 3044
Db 3014 ACAAGCTCAAGCCCTATCTCTTGAGAGCGCAAAAAAGAACCACTTAATTTTAA 3073
QY 3045 AAGTACGTATCACTGTTCTGGAACAGCAAGCATGTGTGTTTGAAGAGTTCAAG 3104
Db 3074 AATATGTATCATCAGCTCCTGGAACAGCAAGCATGTGTGTTTGAAGAGTTCAAG 3133
QY 3105 TCAGAGGGTGGTTCCTCCCAAGCTTACGTGAACCTATTTCCAGGGCCGCTAAGAAATC 3164
Db 3134 TCAGAAAGGGTGGTTCCTCCCAAGCTTACGTGAACCTATTTCCAGGGCCGCTAAGAAATC 3193
QY 3165 CACACATCATGATCTGAGCTCTGAAAGTCTGCTAGTCTAAGAGAGTGGCTCCCG 3224
Db 3194 TACACATCATGATCTGAGCTCTGAAAGTCTGCTAGTCTAAGAGAGTGGCTCCCG 3253
QY 3225 GGGCGCAAGCCAGCCATTCCTCGGAAGAGTTTATTCCTAGTCTAAGAGAGTGGCTCCCG 3284
Db 3254 AGCAGCCAGCCAGCCATTCCTCGGAAGAGTTTATTCCTAGTCTAAGAGAGTGGCTCCCG 3313
QY 3285 TAGAGAGAGAGATTTAACTTTTACAGAGAGAGTGTATGTGTATCCAGAAATAG 3344
Db 3314 TAGAGAGAGAGATTTAACTTTTACAGAGAGAGTGTATGTGTATCCAGAAATAG 3373
QY 3345 TGACTGTGAGACGGGAACGGTGGGCGACAAAGTCCGAGTCTTCCCTTCACTATGTAG 3404
Db 3374 TGACTGTGAGACGGGAACGGTGGGCGACAAAGTCTTCCCTTCACTATGTAG 3433
QY 3405 GCTTAAAGATTAGAGGGCTCTGGAACCTGTGGGAAACAGGAGTTTGAAGAAATAC 3464
Db 3434 GCTTAAAGATTAGAGGGCTCTGGAACCTGTGGGAAACAGGAGTTTGAAGAAATAC 3493
QY 3465 TGAATATTCGCGAGTATTTGCTTCCAGCTGCTACTGCTCCGAAACACTCCCTGGC 3524
Db 3494 TGAATATTCGCGAGTATTTGCTTCCAGCTGCTACTGCTCCGAAACACTCCCTGGC 3553
QY 3525 TCTGTGGGAGCTGATCTGATCCGAGAAAGAACCCAGGTGATGGTGGGAGAGAACT 3584
Db 3554 CCTGTGAGCTGATTTGATCCGAGAAAGAACCCAGGTGATGGTGGGAGAGAACT 3613
QY 3585 GCAAGCTGAGGAGAAAGGCGCAGATAGGGTGGTTCCAGCAATTAATGTCAACTCT 3644
Db 3614 GCAAGCTGAGGAGAAAGGCGCAGATAGGGTGGTTCCAGCAATTAATGTCAACTCT 3673
QY 3645 AAGCCCGGAGAACAGAAATATCCCACTGAGCTAACAGACCGCAGTGCAGCAGC 3704
Db 3674 AAGCCCGGAGAACAGAAATATCCCACTGAGCTAACAGACCGCAGTGCAGCAGC 3733
QY 3705 AGTGTCCAGGTGATCGGAGTGTAGATTACACCGCCAGAAAGATGAGAACTAGCCT 3764
Db 3734 AGTGTCCAGGTGATCGGAGTGTAGATTACACCGCCAGAAAGATGAGAACTAGCCT 3793
QY 3765 CAGCAAGGCGCAGATCATCAAGTCTCTCAACAGAGAGACCCGAGCTGTGTGAAGAGA 3824
Db 3794 CAGCAAGGCGCAGATCATCAAGTCTCTCAACAGAGAGACCCGAGCTGTGTGAAGAGA 3853
QY 3825 AGTCACTGGGAGAGTGGGCTCTTCCATTCATTAATGTAAAGTGAACAAGACATGA 3884
Db 3854 AGTCACTGGGAGAGTGGGCTCTTCCATTCATTAATGTAAAGTGAACAAGACATGA 3913
QY 3885 CCCAGCCAGCAATGATCATATGTTGTGCATCCCGCCAGGGCTTGAAGTCCCTCAA 3944
Db 3914 CCCAGCCAGCAATGATCATATGTTGTGCATCCCGCCAGGGCTTGAAGTCCCTCAA 3973
QY 3945 GAGACCCACTATCCATATCACTGCCAGAGGGATGAGAGATCAGCCTTATCATG 4004

Db 3974 GAGACCCACTATCCCATATCACTGCCAGAGGATGATGAGAGATGACAGCTTATCATG 4033
QY 4005 TGACTTGAGCATGATCACTACTGCTTGTAGTGAAGAACTCACTGCAAGAGCTTT 4064
Db 4034 TGACTTGAGCATGATCACTACTGCTTGTAGTGAAGAACTCACTGCAAGAGCTTT 4093
QY 4065 ACCCTATTTGACCTTATGTCATGTATGAATGTCTGAGTCACTGCGTGCAGAGCGAG 4124
Db 4094 ACCCTATTTGACCTTATGTCATGTATGAATGTCTGAGTCACTGCGAGAGATGAG 4153
QY 4125 AAGC--AAATTCAGAACTGCAAGGGTGGGCTCTTGGGGCTTTCCTAGTCACTC 4182
Db 4154 GAGCAAAATTTACAAAAACACAGGGTGTGGGCTCTTGTGGCTTCTAGTCACTC 4213
QY 4183 AAGCTGACCGG--CCCGGCTTCAACAGGGCGCTTCAATAGTTTAAATTAATTTTAA 4241
Db 4214 AATATGACTTTCCTCCACCTTTCACAGGGTGTTCCAATAGTTTAAATTAATTTTAA 4273
QY 4242 TGTATTTTACCTTTTATTAATAATCTCAATCACTCTTGGCTATTTGGTTT 4301
Db 4274 TATATTTTATGCTTTTATTAATAATAATAATTAATCACTTGTGCTATTTGGTTT 4333
QY 4302 TACAAAAACACCATATCAAGAGTGCCTGTGCGAGCATTAATAATGCTGTTCCGG 4361
Db 4334 TCAAAAAAGACCATATCAAGAAATG----CTGATGTGCTATTAATAATGTTCCAA 4388
QY 4362 CGTACCGTAACTGAGAGCTGCTGTACTTTGGCGGTGTGCTGAGTTCCTCAACAT 4421
Db 4389 ATGTCAATTAATCTGAGACTTATGTAATTTTTCATTTGTGCTGAGTTCCTCAACAT 4448
QY 4422 TGTATGTTTGGGGCTGTTCC--TGCCTAGAGCAGAGAGATGGGTGATCTGT 4477
Db 4449 TGTGAGTTTGGGGCTTTCCTTCCATGAAAGTGCAGAGAGTGTAGATCTGT 4508
QY 4478 TTTGAATATGTATGTAGACTGAGCTGACTATGAAAGGGTATGCTGTCTGTAGC 4537
Db 4509 TTTGAATATGTATG--GAATGAGCCCAATTAAGCGAAGTGTGTGCTGTGTGT 4567
QY 4538 ATCAGGTGAC--TGTGCGGATGTACATCTGTACCGAAGATGACT--CTTCTCAT 4595
Db 4568 ATCAGGTGACCTTGTGAGCATGTATACATCTGTATGAAGAAATGATCTTTCCA 4627
QY 4596 GGTAAACCCACACCGGTGACAGTCTCTCATCT--GCATCATTTTCTTTGAC 4653
Db 4628 TGCAGAGCTTATCTGTGATGATGCTCTATCAATTTGATTTATTTTGGAC 4687
QY 4654 AGTGACTTGTAGCAGCTGAGAGAC--CCATGTTCCGTTGGTCTCAGATGACT 4712
Db 4688 AGTGACTTGTAGCAGCTGAGAGACCTGCTGTTTGTGTGCTCAGATTTATCT 4747
QY 4713 AGTGTCCCGTGTGTTTATTTTCAATCTGCAATGTCTTCAACATTAACCTAG 4772
Db 4748 GGTATAGTTGGTGTGTTTGTGGGTTTATTTTGGCTGTTCATGATTAATAATCA 4807
QY 4773 TAAAGCGCACTGCGCGGGTTACATCATCATCCACCGTC-----TTAGTCTC 4827
Db 4808 GTAGACAC--ACCATGAGGTGTATCAGATCAACATATCCACAGCTCTTTTATGCTC 4866
QY 4828 TGTACGTGAG--TTATTTCCAGTGTCTTTATGAA-----TATCTTGAACAATAT 4881
Db 4867 TGTATGATGATTTTATTTCCAGTTACTTTTATGGAATGACCTTTTGAACAATAT 4926
QY 4882 CTTCTTGAAGAAAGATGTATGAGTCTTCTGCAATTAATTTCCAGTGTTCAT 4941
Db 4927 TTTCTTGAAGAAAGATGTATGAGTCTTCTGCAATTAATTT--CCAAATTTTACAT 4985
QY 4942 TTTTAACTAGCTGTGGGGTGTGCTACAGATTAATGAAGTGGCGTCCGTGCGG 5001
Db 4986 TTTTAACTAGCTGT--GAATTTCTACAGATTAATGAAGTGGCGTCCGTGCGG 5044
QY 5002 TGTGTATTAATCTTGTGCTGATGAGCGGTGTCT--CTTATGATATTAATGAGT 5058
Db 5045 TGTGTGT--AGATATGCTGTGATGAGCGGTGTCTTCTTTAAACCTATGTTGAAGC 5103

CC 5059 CGGAGAGAAATT 5071
 CC |||||
 DB 5104 TCTCATATAAAT 5116

RESULT 6
 ID AA234572 standard: cDNA, 5195 BP.
 AC AA234572;
 DT 01-FEB-2000 (first entry)
 XX Human SH3D1A cDNA clone 11.
 DE
 XX SH3D1A gene; human; Down's syndrome; leukaemia; cancer;
 KW megakaryocytic abnormality; myeloproliferative disorder;
 KW platelet disorder; neural disorder; thrombocytopenia;
 KW haematopoietic disorder; cognitive dysfunction; microcephaly;
 KW lissencephaly; colpocephaly; holoprosencephaly; diagnosis; therapy;
 KM 88.
 XX Homo sapiens.
 OS
 FH Key Location/Qualifiers
 FT CDS 239..3886
 FT /*tag= a
 XX MO9953062-A2.
 XX 21-OCT-1999.
 XX 16-APR-1999; 99WO-US08371.
 XX PR 16-APR-1998; 98US-0082007.
 XX (CEDA-) CEDARS SINAI HEALTH SYSTEM.
 PA Korenberg JR, Chen X;
 PI MPI; 1999-633829/54.
 DR P-PSDB; AAY32156.
 XX
 PT Nucleic acid from the human SH3D1A gene and its products, useful for
 PT the diagnosis and treatment of myeloproliferative disorders and
 PT leukaemia
 PS
 PS Claim 2; Fig 10; 99p; English.
 XX This is the nucleotide sequence of full-length cDNA (clone 11)
 CC corresponding to a novel human SH3 gene, termed the SH3D1A gene,
 CC that contributes to the development of platelets and the
 CC pathogenesis of leukaemias, both in general and in particular those
 CC involving the megakaryocytic lineage. The SH3D1A gene maps to the
 CC small candidate region for low platelets on chromosome 21.
 CC Sequencing of 5 different sizes of cDNA clone from foetal brain
 CC (see AA234570-74) suggests that at least 3 isoforms exist. The
 CC invention provides methods for the diagnosis and treatment of
 CC megakaryocytic abnormality, myeloproliferative disorder, platelet
 CC disorder, acute leukaemia, neural disorders, thrombocytopenia,
 CC platelet disorder on chromosome 21, low platelets in deletion for
 CC 21, association of gains in chromosome 21 with leukaemias, neural
 CC abnormalities, dysfunctions and disorders including brain
 CC malformations and corresponding cognitive dysfunctions,
 CC microcephaly, lissencephaly, and colpocephaly. Methods are also
 CC provided for: suppressing cells unable to regulate themselves;
 CC screening for a somatic alteration in the SH3D1A gene; monitoring
 CC the progress and adequacy of a treatment; monitoring tumour risk
 CC progress or megakaryocytic abnormality, myeloproliferative disorder,
 CC haematopoietic disorder, platelet disorder or leukaemia; and
 CC treatment of a subject (including a prenatal subject) having
 CC megakaryocytic abnormality, myeloproliferative disorder,

CC platelet disorder, leukaemia or neural disorder using a
 CC nucleic acid that expresses SH3D1A or its antisense nucleic acid.
 XX
 SQ . Sequence 5195 BP; 1592 A; 1148 C; 1237 G; 1218 T; 0 other;
 Query Match 61.8%; Score 3141.8; DB 20; Length 5195;
 Best Local Similarity 80.1%; Pred. No. 0;
 Matches 4069; Conservative 0; Mismatches 742; Indels 272; Gaps 21;

QY 20 GCGGCGGAGAGGCGGCGGAGCTTGTCTCCGTAGTACGGCGGCTTGGCAAGGAGCAT 79
 DB 10 GCGCCAGGAGGAGGAGGCTTGTGTCTCCGTAGTACGGCGGCTTGGCAAGGAGCAT 69
 QY 80 CCCGAGCGGAGCTCCGGGACGGCCGGAGGAGGAGGCGGCGGAGTGTGTGCG 139
 DB 70 CCCGAGCGGAGCTCCGGGAGC-----GACGAGAGGCGGCGGAGTGTGTGCG 118
 QY 140 CGGCTGGGAGCTCGGCGCTCTCTGC-GCGGCGTGGGCTGCACTGATTGTGTGAGGG 198
 DB 119 GGGCTGGGAGCTCTGCGTCCCTCCAGCGGCGGCTGAGCGGCACTGATTGTCTCGGG 178
 QY 199 CGGCGGCGGAGCGGCGGCGGAGTGAAGCGTGCATGACGAGGTGAACGTATGAAAC 258
 DB 179 CGGCGGCGGAGCGGCGGCGGAGTGAAGCGTGCATGAGCAAGGTGAAGTGAAGCAAC 238
 QY 259 ATGAGCTAGTTCCACACCTTGGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 318
 DB 239 ATGAGCTAGTTCCACACCTTGGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 298
 QY 319 GAAAGGCGCAAGCATGACGAGCTTCTAGCTGAGGCGGAGTGAAGTGAAGTGAAGTGAAG 378
 DB 299 GAAAGGCGCAAGCATGACGAGCTTCTAGCTGAGGCGGAGTGAAGTGAAGTGAAGTGAAG 358
 QY 379 GGTGATCAAGGAGAACTTTTTCATCTGGGTTAAGCTGAGCTGTCTTGACAA 438
 DB 359 GGTGATCAAGTGAAGAACTTTTTCATCTGGGTTAAGCTGAGCTGTCTTGACAA 418
 QY 439 ATATGGGCGCTTGAAGGAGCATGATGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 498
 DB 419 ATATGGGCGCTTGAAGGAGCATGATGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 478
 QY 499 GCCATGAAGCTTATCAACTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 558
 DB 479 GCTATGAAGCTTATCAACTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 538
 QY 559 GTCATGAAGCAAGCAAGTGTATTTCCAGTGCACAGATTTGGTATGAGGAGATT 618
 DB 539 GTCATGAAGCAAGCAAGTGTATTTCCAGTGCACAGATTTGGTATGAGGAGATT 598
 QY 619 GCTAGCATGACCACTGACAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 678
 DB 599 GCTAGCATGACCACTGACAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 658
 QY 679 GAAATGTCTCCACCTTATGATCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 738
 DB 659 GAAATGTCTCCACCTTATGATCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 718
 QY 739 GGGGCTCTCCCTGATACAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 798
 DB 719 GGGGCTCTCCCTGATACAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 778
 QY 799 AAGAGTTCTTCTTACAGAGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 858
 DB 779 AAGAGTTCTTCTTACAGAGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 838
 QY 859 GCACATCATTTGATGTGCGCAGGCGCTTCCAGCAGCAGATGAGTGGCTGTCTCAGTCA 918
 DB 839 GCACATCATTTGATGTGCGCAGGCGCTTCCAGCAGCAGATGAGTGGCTGTCTCAGTCA 898
 QY 919 TCAAGCTGAATACAGGCACTTATTCACAGCAGCAAACTATGATGAGTGAACACTTA 978
 DB 899 TCAAGCTGAATACAGGCACTTATTCACAGCAGCAAACTATGATGAGTGAACACTTA 958

Db 3119 GTGAACCTCATTTCAGGGCCCATAGAAAGTCTACAGAGATGATTTCTGTTCTTCAGAG 3178
 Qy 3193 AGTCTGTAGTCTTAAAGAGAGTGGCTTCCTCCGGCCGCAAGCCCATTCCTCCGAGAA 3252
 Db 3179 AGTCTGTAGTCTTAAAGAGAGTGGCTTCCTCCAGAGCCAGCCGCTGTTCCGAGAA 3238
 Qy 3253 GAGTTTATTCGATGTACATACAGAGAGTTCGAGCAAGAGATTTAACTTTTCAGCA 3312
 Db 3229 GAATTTATTCGATGTACATACAGAGAGTTCGAGCAAGAGATTTAACTTTTCAGCA 3298
 Qy 3313 GGGGATGTATTTGTTGTTACCAAGAAAGTGTGACTGTGTGACGAGAAACGCTGGGCA 3372
 Db 3299 GGGGATGTATTTGTTGTTACCAAGAAAGTGTGACTGTGTGACGAGAAACGCTGGGCA 3358
 Qy 3373 AAGTCCGAGTCTTCCCTTCTTACATATGTAGAGCTTTAAAGATTCAGAGGCTTCGAACT 3432
 Db 3359 AAGGCGGAGTCTTCCCTTCTTACATATGTAGAGCTTTAAAGATTCAGAGGCTTCGAACT 3418
 Qy 3433 GCTGGGAAAAACAGGAGTTTAAAGAAAAAACTGAAATTTGCCAGGTTATTGCTCTAC 3492
 Db 3419 GCTGGGAAAAACAGGAGTTTAAAGAAAAAACTGAAATTTGCCAGGTTATTGCTCTAC 3478
 Qy 3493 GCTGCTACTGTCTCCGAAACAACCTGCTGCTGCGCAGCTGATTTCTGATCCGAA 3552
 Db 3479 ACCGCAACCGGCCCCGAGCAGCTCACTCTGCGCTGCTGATTTGATCCGAA 3538
 Qy 3553 AAGAACCAAGGTGTATGTGTGGAAAGAGAACTGCAAGCTCCAGGAAAAAGCCAGTA 3612
 Db 3539 AAGAACCAAGGTGTATGTGTGGAAAGAGAGCTGCAAGACGTTGGAAAAAGCCAGTA 3598
 Qy 3613 GGGTGGTTTCCAGCAAAATTTATGTCAAACTTTCAACCCCGCAAGCAAAATCACTCCA 3672
 Db 3599 GGGTGGTTTCCAGCAAAATTTATGTCAAACTTTCAACCCCGCAAGCAAAATCACTCCA 3658
 Qy 3673 ACTGAGTACCCAGAACCGCAGTGCAGCAGAGTGTGCGCAGTGTATGAGTACAT 3732
 Db 3659 ACAGAGCACTTAAGTCAACAGCATTAGCGGAGTGTGCGCAGTGTATGAGTACAT 3718
 Qy 3733 TACACCGCCAGAAAGATGACCACTAGCTTACCAAAAGGCAATATCAACCTCTC 3792
 Db 3719 TACACCGCCAGAAAGATGACCACTAGCTTACCAAAAGGCAATATCAACCTCTC 3778
 Qy 3793 AACAGAGAGAACCGCAGTGTGTGAAAGAGAGTCAAGTGGGCAAGTTGGCTTTCCA 3852
 Db 3779 AACAGAGAGAACCTGAGTGTGTGAAAGAGAGTCAAGTGGGCTTTCCA 3838
 Qy 3853 TCCAAATATGTAAAGCTGACCAAGACATGAGCCCGCAGCAGCAATGATATGTTGT 3912
 Db 3839 TCCAAATATGTAAAGCTGACCAAGACATGAGCCCGCAGCAGCAATGATATGTTGT 3898
 Qy 3913 CCATCCCCCTCAGAGCTTTGAAATCTCTAAAGAACCCCATATCCCATATCACTGCCA 3972
 Db 3899 CCATCCCCCTCAGAGCTTTGAAATCTCTAAAGAACCCCATATCCCATATCACTGCCA 3921
 Qy 3973 GAGGATGTATGAGATGAGCCTTGATCATGTAGTCTGACAGCATGATCACTACTGCT 4032
 Db 3922 ----- 3921
 Qy 4033 TCTGATGAGAAAGTCACTGACAGAGAGTTTAACTTGAACCTTAGTGCATGTAT 4092
 Db 3922 ----- 3921
 Qy 4093 CGAAATGTCTGATGACTGCGTGCAGAGGAGAGCAAAATTTGCAGAACTGCAGAGGTGG 4152
 Db 3922 ----- 3921
 Qy 4153 TGGGTCCTTTTGGGCTTTCTCTAGTCACTCAGACTGAC-CGGCCCGCTTCACAGGGC 4211
 Db 3922 ---GTCCTTTTGGGCTTTCTCTAGTCACTCAGACTGACCTTTCCACAGAGT 3978
 Qy 4212 GCTTTCATATGATTTTAAAGATTTTAAATGTGTATTTTAACTTTTAAATAAATCTC 4271
 Db 3979 GCTTTCATATGATTTTAAAGATTTTAAATGTGTATTTTAACTTTTAAATAAAT 4038

Qy 4272 AATCAATTAATCTTCTTGGCTATTTTGGTTTACAAAAACACCACTAATCAAGAGTGGCT 4331
 Db 4039 AATTAATTAATCTTCTTGGCTATTTTGGTTTACAAAAACACCACTAATCAAGAGTGGCT 4095
 Qy 4332 GTCGCGGACGATTTAAATGCTGTTCGCGGCGTACCGTAACTGAGAGCTTGTCTGACT 4391
 Db 4096 --CTGCAATGTCTATTTAAATTTGTTCCAAATGTCCATTAATCTGAGACTGTATTT 4153
 Qy 4392 TTGCGCTTTTGTCCAGTGTTCCTCAACACATTTGTGTATTTTGGGCTGTTCCTC---TCCC 4447
 Db 4154 TTTCAATTTTGTCCAGTGTTCCTCAACATTAATTTGTGACGTTTGGGCTTTTCCCTTACA 4213
 Qy 4448 GTAGAGCAGAGAGATGAGTGTACCTGTTTGAATATGTATGTAGACTGAGCCCTGA 4507
 Db 4214 TAGAATGTCAAGAGATTTCAATATCTGTTTAAAGAGTATGAATGAGCCCAATTA 4273
 Qy 4508 CTATGAAAGGGGTTATGCTGTGTGTGACCATCACTGATACC-TGTGCGCATGTACCAT 4566
 Db 4274 AGC-GAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4332
 Qy 4567 CTGACCGAAGATGACT-CTTCTCTCATGCTTAAACCAACCGTGTACAGTGTCT 4625
 Db 4333 ATCTGTACATTAAGAAATTAATGTTCTTTCATGCGCAAGCTATTAATCTGTACAGTGTCT 4392
 Qy 4626 CATCTACT--GCATTCATTTTACTTTCGACAGTACCTGTGAGCCCTGAGAGAGAC 4683
 Db 4393 AATCATATTTGATTTAAATTTTATTTTGTGACAGTACCTTTGAGCAGCATGAAAGACT 4452
 Qy 4684 CA-TGTTTCCGTTTGTGTCTGATGTACCTGATGTGCCCCGTTTGTTTTATTTTTC 4742
 Db 4453 CTGTGTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4512
 Qy 4743 AATCTGCAATCTTTCACACATTAATTAAGTAAAGCCCACTGCGGCTTACAT 4802
 Db 4513 AATTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4571
 Qy 4803 CATCAGTACCCAGCGTC-----TTAGTCTGTGTACGGAAG-TTATTCAGTGTCTT 4856
 Db 4572 CACGATATTCACAGTCTCTTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4631
 Qy 4857 TTATGAA-----TATCTGAAACAGTATCTTCTTCAAGAAAGATGTATGAAATC 4911
 Db 4632 TCATGAAATGACATATTTTGAACAGTAAATTTCTTGAAGAAAGATGTATGAAATC 4691
 Qy 4912 TCCCTGCAATTAATTTCCAGTGTTCATATTTTAACTTACAGCTGTGGGCTGTACAG 4971
 Db 4692 TCCCTGCAATTAATTTTCCAGTGTTCATATTTTAACTTACAGCTGTGGGCTGTACAG 4749
 Qy 4972 ATTAATATGAAATGAGCGCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5031
 Db 4750 ATTAATATGAAATGAGCGCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4808
 Qy 5032 GTGTGT---CCTTAGATATTTAGTGAAGTGTGGAAGGATTT 5071
 Db 4809 CTGTTTGTCTTTAAACATAGTTTGAAGCTCTCATTAATAAT 4851

RESULT 7
 AA234570
 ID AA234570 standard; cdna; 5199 BP.
 XX AA234570;
 AC
 AC
 DT 01-FEB-2000 (first entry)
 DE Human SH3D1A cDNA.
 XX
 XX
 XX SH3D1A gene; human; Down's syndrome; leukaemia; cancer;
 KM megakaryocytic abnormality; myeloproliferative disorder;
 KM platelet disorder; neural dysfunction; thrombocytopenia;
 KM haematopoietic disorder; cognitive dysfunction; microcephaly;
 KM lissencephaly; colpocephaly; holoprosencephaly; diagnosis; therapy.

KW ss.
 XX Homo sapiens.
 OS
 XX
 XX
 FT Key Location/Qualifiers
 FT CDS 208..3642
 FT /*tag= a
 PN MO9953062-A2.
 XX
 XX
 PD 21-OCT-1999.
 XX
 PF 16-APR-1999; 99WO-US08371.
 XX
 PR 16-APR-1998; 98US-0082007.
 XX
 PA (CEDA-) CEDARS SINAI HEALTH SYSTEM.
 XX
 PI Korenberg JR, Chen X;
 XX
 PS WPI: 1999-633829/54.
 DR P-PSDB; AA32154.
 XX
 PT Nucleic acid from the human SH3D1A gene and its products, useful for
 PT the diagnosis and treatment of myeloproliferative disorders and
 PT leukaemia -
 XX
 PS Claim 2; Fig 5; 99pp; English.
 XX
 XX This is the nucleotide sequence of full-length cDNA corresponding
 CC to a novel human SH3 gene, termed the SH3D1A gene, that contributes
 CC to the development of platelets and the pathogenesis of leukaemias,
 CC both in general and in particular those involving the small
 CC megakaryocytic lineage. The SH3D1A gene maps to the small
 CC candidate region for low platelets on chromosome 21. Sequencing
 CC of 5 different sizes of cDNA clone (see AA324570-74) suggests that
 CC at least 3 isoforms exist. The invention provides methods for the
 CC diagnosis and treatment of megakaryocytic abnormality,
 CC myeloproliferative disorder, platelet disorder, acute leukaemia,
 CC neural disorders, thrombocytopenia, platelet disorder on
 CC chromosome 21, low platelets in deletion for 21, association of
 CC gains in chromosome 21 with leukaemias, neural abnormalities,
 CC dysfunctions and disorders including brain malformations and
 CC corresponding cognitive dysfunctions, microcephaly, lissencephaly,
 CC and colpocephaly. Methods are also provided for: suppressing
 CC cells unable to regulate themselves; screening for a somatic
 CC alteration in the SH3D1A gene; monitoring the progress and
 CC adequacy of a treatment; monitoring tumour risk progress or
 CC megakaryocytic abnormality; myeloproliferative disorder.
 CC haematopoietic disorder, platelet disorder or leukaemia; and
 CC treatment of a subject (including a prenatal subject) having
 CC megakaryocytic abnormality, myeloproliferative disorder,
 CC platelet disorder, leukaemia or neural disorder using a
 CC nucleic acid that expresses SH3D1A or its antisense nucleic acid.
 XX
 SO Sequence 5199 BP; 1615 A; 1152 C; 1214 G; 1218 T; 0 other;
 Query Match 60.6%; Score 3080.2; DB 20; Length 5199;
 Best Local Similarity 79.8%; Pred. No. 0;
 Matches 4030; Conservative 0; Mismatches 753; Indels 268; Gaps 23;

QY 235 CAGCAGGTGAAGTATAGTAAACCATGGCTGATTTCCACACCTTTGGTGTAGCTG 294
 DB 184 TAGCAAGGTAAAGTAAAGTAAAGCAATGCTGCTGCTTCCAAACCTTTGGTGTAGCTG 243
 QY 295 GATGTCTGGGACCTAACTGTGTGAGAGAGAGGCGCAAGCATGACGAGTCTTACCTG 354
 DB 244 GATATCTGGGACCTAACTGTGTGAGAGAGAGGCGCAAGCATGATCAGAGTCTTACCTG 303
 QY 355 AAGCCATGACGGGAGTATTTACTGTGTATCAAGCAGAGAACTTTTTCATCTGGG 414
 DB 304 AAGCCATGATCTGGATTTACTGTGTATCAAGCAGAGAACTTTTTCATCTGGG 363
 QY 415 TTACCTCAGCCTGTCTTACACAAATATGGGCGCTGAGGAGATGAATACATGAGAG 474
 DB 364 TTACCTCAGCCTGTCTTACACAAATATGGGCGCTGAGGAGATGAATATGATGAGAG 423
 QY 475 ATGGATCAAGTGAATTTTTCATAGCAGTAAAGCTTATCAAACTGAGTACAGAGATAT 534
 DB 424 ATGGATCAAGTGAATTTTTCATAGCAGTAAAGCTTATCAAACTGAGTACAGAGATAT 483
 QY 535 CAGCTCCCTCCACACTTCCCTCTGATGAAACAGCAACAGTGGCTATTTCCAGTGA 594
 DB 484 CAGCTACCTCTGCACTTCCCTCTGATGAAACAGCAACAGTGGCTATTTCCAGTGA 543
 QY 595 CCAGCATTTGATAGGAGGATTTGCTAGATGCCACCTGACAGCTGTGCTCTG 654
 DB 544 CCAGCATTTGATAGGAGGATTTGCTAGATGCCACCTGACAGCTGTGCTCTG 603
 QY 655 CCAATGGCTCAATTCAGTTGTAATGTCTCCACCTTATATCTTCTGCTCTCA 714
 DB 604 CCAATGGATTCATTCAGTTGTAATGTCTCCACCTTATATCTTCTGCTCTCA 663
 QY 715 GCAAGCATGCTCTCCCTGCTTAAAGGCTCTCTCTGATACAGCTCTGCTGCTT 774
 DB 664 GCAAGCATGCTCTCCCTGCTTAAAGGCTCTCTCTGATACAGCTCTGCTGCTT 723
 QY 775 GGGCATCTGAGGCCACATGGCCAAAGATTTCTCTTGAAGATCTGTCCAGGCTCA 834
 DB 724 GCTCATCTGAGGCCACATGGCCAAAGATTTCTCTTGAAGATCTGTCCAGGCTCA 783
 QY 835 CAATTAACTAATGTTACAGAGGACATATTCATGATGCGCCAGCCCTCCAGCA 894
 DB 784 CAACTAACTAATGTTACAGAGGACATATTCATGATGCGCCAGCCCTCCAGCA 843
 QY 895 GCAAGATGGCTGTGCTGCTATGATCAGCTGAAATACAGCAGTATTCACAGCCAC 954
 DB 844 GCAAGATGGCTGTGCTGCTATGATCAGCTGAAATACAGCAGTATTCACAGCCAC 903
 QY 955 GCAAAATCTAATGTTACAGAGGACATATTCATGATGCGCCAGGAAATTTCTCAATCA 1014
 DB 904 GCAAAATCTAATGTTACAGAGGACATATTCATGATGCGCCAGGAAATTTCTCAATCA 963
 QY 1015 AGTTTACCCAGGCTCAGCTGCTTCAATATGAAATCTTTCGACATTTGATCAAGATGA 1074
 DB 964 AGTTTACCCAGGCTCAGCTGCTTCAATATGAAATCTTTCGACATTTGATCAAGATGA 1023
 QY 1075 AAATCTACTGCAAGAAATTTATCTAGCTATGACCTAATGATGTTCCATGTCTGT 1134
 DB 1024 AAATCTACTGCAAGAAATTTATCTAGCTATGACCTAATGATGTTCTGTCTGTG 1083
 QY 1135 CAGCAGTGGGCGGCTGCTGCTCAGAAATCATCTCCCTTCTTCAAGAGAGTTGCG 1194
 DB 1084 CAGCAGTGGGCGGCTGCTGCTCAGAAATCATCTCCCTTCTTCAAGAGAGTTGCG 1143
 QY 1195 TCCGAGTGGGATGCTGCTTAAAGCTTTCTTGTGTGATCAGAGCTGCTGAGAG 1254
 DB 1144 TCCGAGTGGGATGCTGCTTAAAGCTTTCTTGTGTGATCAGAGCTGCTGAGAG 1203
 QY 1255 CCGTGTGAGAGATGAGCAGAC---CAGAGAAAGAACTCCCTGTGACATTTGAAGAT 1311
 DB 1204 CCGTGTGAGAGATGAGCAGAC---CAGAGAAAGAACTCCCTGTGACATTTGAAGAT 1263
 QY 1312 AAGAGCGGGAACCTTGAAGGAGGAGTGTGAGCTGGAAGAGCGCCAGAGGCTC 1371

Db	1264	AAAGAGCCGGAGAACTTTTGAACGTGTGGCAACTGGAACTGGAGAAACGAAGCCAACTCTC	1323
Oy	1372	TTGGAGCAGACAGCCGCAAGAGCAGAGCGGTTGGCTCAGCTGGAGACGGCCGAGCAGAG	1431
Db	1374	CTGGAAACAGCAGCCGAGAGAGAGAGCGCTGGCCCAAGCTGTGAGCCGGGCGAGACAGAG	1383
Oy	1432	AGAAAGAGCCGGAGCCCGCAGAGCAGAGAGCCCAAGCGCAGCTGGAGCTGGAGACAG	1491
Db	1384	AGGAAGAGCGTGAACGCCAGAGCAAGAGCCCAAAACAACTGGAACCTGGAGAAACGA	1443
Oy	1492	CTGGAGAAAGCAGCCGGAGCTGGAGGGCAGCAGAGAGAGAGAGAGAGAGAGATCCAG	1551
Db	1444	CTGGAAAAGCAGCCGGAGCTAAGACGCAAGAGAGAGAGAGAGAGAGAAAGAAATTGAG	1503
Oy	1552	AGGCGCAGAGCCGCAAAAACGGGAACCTGAAAAGCAGACCAACTTGAATGGGAACGGAAC	1611
Db	1504	AGGCGAGAGGCTGCAAAACGGGAACCTTGAAGGCAACGCAACTTGAATGGGAACGGAAT	1563
Oy	1612	CGAGACACAGAACCTCTGTAATCAGAGAACAGAGCAGAGAGGACCGTGGTCTTGAG	1671
Db	1564	CGAAGGCAGAAACTACTTAATCAAAAGAAACAAAGAGCATGTGTTGTACTGAAA	1623
Oy	1672	GCAAGAGAGAAAGACTCTGTGAGTTTGAAGTTGAAGCTCTGAATGAACAAAGATAGCTA	1731
Db	1624	GCAAGAGAAAGACTCTTGGAAATTGGAATTGAAAGCTCTAAATGAATAAAGATCACTA	1683
Oy	1732	GAGGAAAACTTCAGAGATATCAGGTGTGCACTGGGCAACCCAGAGGCAAAATTGAGAC	1791
Db	1664	GAGGGAATCTTCAGAGATATCAGATGTGATTTGACACCCCAAGGCAAAAGATTTAGAGC	1744
Oy	1792	ACGAACAAAGCTTGAAGAGCTTGAAGATTGCTGAATCACCCTTACAGCAGCAGTTGAG	1851
Db	1744	ACAAACAATCTAGAGAGTTGAGATTGGCCGAATCCACCCTACACCAACTTACG	1803
Oy	1852	GAATCTCAGCAAAATGCTTGGAAAGACTTATTCAGAGAAACAGATCTCAGTGAACAGTTA	1911
Db	1804	GAATCTCAGCAAAATGCTTGGAAAGACTTATTCAGAGAAACAGATCTCAGTGAACCAATTA	1863
Oy	1912	AAACAGATCCAGCAGAAACAGTTTGGATAGAGACTGCTTCTTAACCTCAAAAAGACCTTG	1971
Db	1864	AAACAGATCCAGCAGAAACAGTTTGGACAGAGATTAACCTTTTACCTTAAAAAGACCTTA	1923
Oy	1972	GAGCAAAAGAGAGTGGCCCGCAGCAGCTCCGGAGCAGCTGGACAGAGGTGAGAGAG	2031
Db	1924	GAGCAAAAGAGACTGAGCTCCGAGCAGACCTACAGAGCAACTGGATGAAGTGAAGAAAG	1983
Oy	2032	ACCAGGTCAAAGCTGCAGAGATTTATGTTTTCAACAACAGCTGAAGAACTGAGAG	2091
Db	1984	ACTAATATCAAAACTTCAGAGAGATTATATTTTCAATAATCAGCTAAAGAACTAAGAGAA	2043
Oy	2092	ATACATACCAAAACACCACTCCAGAAAGCAGAGGTCCCTGGAGGCGCGACTGAAGAG	2151
Db	2044	ATACCAATTAAGCAACAACTCCAGAAAGCAAAAGTCAAGAGGCTGAACGACTGAACAG	2103
Oy	2152	AAAGAGCAGAGAGAGAGGCTCTGAGTTAGAGAAACAAAGAGAGAGCGCTCAGAGAG	2211
Db	2104	AAAGAAACAGAACGAAAGATCATTAATTAAGAAAAACAAAAGAGAGGCCCAAGAGAG	2163
Oy	2212	GTTTCAGAGAAAGGACAAAGCAATGCTTGAAGCATGTGCAGCAGAGAG--GAGCAGCCAGC	2266
Db	2164	GCTCAGAGAAAGGACAAAGAGTGGCTGGAGCATGTGCAGCAGAGAGAGAGAGCATCAGAG	2222
Oy	2266	CCCCGGAACCCCAAGAGAGAGACAGCTGAAGAGAGGAAGCAGTGCAGGAAGAAAGAG	2322
Db	2224	CCAAAGAAATCTCAAGAGAGGAAAACTGAAAAGAGAGAGAGAGTCAAAAAGAGAGAT	2283
Oy	2339	GCGAGAGAGAGAGCCAGCCGGAATAATGCAGAGCAAGAGAGTCCGCTTTTCCATCCGAT	2388
Db	2284	GCGAGAGAGAAAGGAGAAACAGGAACCAAGAGCAAGAGTGGGTCCGCTTTTCCATCAACAC	2344
Oy	2389	CAGAGCCAGCTTAAGCTGGCCACCCAGAGCACCTGGTGTACCAACAGAGAAAGCCCGCTT	2446

Db	2334	CAGAACCAGTAAAGCCAGCTGTCCAGGCACCCTTGTCCTCACTGCAGAAAAAGTCTCACTT	2403
QY	2449	ACCAATTTCTGCACAGAGAGTGTAAAAAGTGTATATTACCGAGCGCTGTACCCCTTTGAA	2508
Db	2404	ACCAATTTCTGCACAGAAAAATGTAAAAAGTGTGTATTACCGGCGCACTGTACCCCTTTGAA	2463
QY	2509	TCCAGAAAGTCAACGATGAGATTCACATCCAGCCAGAGATATATAGTCAATGTGTGAATGAAGC	2568
Db	2464	TTCCAGAAAGCCATGATGAAATACATATCAGCCAGAGACATATGTCATGTGTGAATGAAGC	2523
QY	2569	CAGACTGAGAGCCAGAGATGAGCTTTGAGAGAGAGCTGAAAAGGAGACAGGAGTGGTTCCCT	2628
Db	2554	CAAACTGGAGAACCCCGCTGGCTTTGAGAGAGAAATTTAAAAGAAAGACAGGTTGGTTCCCT	2583
QY	2629	GCAAACTATGCGAAAAAGATTCAGAAAAATGAGGTTCCCACTCCAGCCAAACCAATGACC	2688
Db	2584	GCAAACTATGCGAGAAAAATCCAGAAAAATGAGGTTCCCGCTCCAGTGAACCACTGACT	2643
QY	2689	GATCTGACATTTGGCCCCCTGGCCCCCAAACTGGCTCTGGGTGAGAACCCCTGCTCTTTGGCA	2748
Db	2644	GATTCAAATCTGGCCCCCTGGCCCCCAAACTGGCTCTGGGTGAGAACCCCGCTTTGGCA	2703
QY	2749	GTTGACCTTTTGAAGCCCCCTCCAGAACCCCAACAACTGGGAGAGCTTCAGTTCCACGTGG	2808
Db	2704	GTTAACTTTTGAAGCCCTCCACGACCCCTTATATATCTGGGCGGACTTCAGCTCCACGTGG	2763
QY	2809	CCCAAGCAGCTTCAAACGAGAACCCAGAAACGACAACTGGGATAGCTGGGCGGCTCAGCT	2868
Db	2764	CCCAACGACACGAAATGAGAAACGAGAAACGGAATATCTGGGATGATGGGCGCCAGCC	2823
QY	2869	TCTCTGACCGTACTTATGTGCTGGCGCAGTTACGCGAGAGATAGGCTTTTACCCGAGCCACA	2928
Db	2824	TCTCTCACCGTTTCCAAAGCCGCGCGCAGTTAAGGCGAGAGGTCGGCTTTTACTCCAGCCAG	2883
QY	2929	GCGACTGGCTCTGCCCATCTCCCGCTCCGGGCGAGGTTGAAAAGTGGAAAGGAGCTACAA	2988
Db	2884	GCGACTGGCTCTCCCGCTCTCTCTGTGTAGGCGAGGTTGAAAAGTGGAGGAGGCTACAA	2943
QY	2989	GCGCAGCGCCCTGTATCCCTGGAGAGCCAAAAAGACAAACCTTTAAATTTTAAACAAAAGT	3048
Db	2944	GCTCAAGCCCTATATCTCTGGAGAGCCAAAAAGACAAACCTTTAAATTTTAAACAAAAT	3003
QY	3049	GACGTCAATCACCGTTCTTGGAAACAGCAAGACATGTGTGTGTTTGGAGAAAGTTCAAGGTACG	3108
Db	3004	GATGTCAATCACCGTCTCTGGAAACAGCAAGACATGTGTGTGTTTGGAGAAAGTTCAAGGTACG	3063
QY	3109	AAAGGTTGGTTCCCAAGTCTTTAGGTGAAATCATCTTTCAGGGCCCGTAAGGAAATTCACA	3168
Db	3064	AAAGGTTGGTTCCCAAGTCTTTAGGTGAAATCATCTTTCAGGGCCCGTAAGGAAATTCACA	3123
QY	3169	AGCATTCATCTGAGCCCTACTGAAAGTCTTGCTAGTCTTAAAGAGAGTGCTTCCCGGCC	3228
Db	3124	AGCATTCATCTGAGTTCTTGATGAGAGAGTCTGTAGTCTTAAAGAGAGTAGCTCTTCCAGCA	3183
QY	3229	GCCAAAGCCAGCCATTTCCCGAGAAAGTTTATTGCCATGTACACATACGAGATTTCTGAG	3288
Db	3184	GCCAAAGCCAGCTGTTCGGAGAA-----	3207
QY	3289	CAAGGAGTTTAACTTTCAACGAAAGGGGATGTGATTTGTGTTTCAAGAAAGATGTGTAC	3348
Db	3208	-----	3207
QY	3349	TGTTGAGACGGAGACGTTGGCGCAAGTCCGAGTCTTCCCTTCTAACTATGTAGGCTT	3408
Db	3208	-----	3207
QY	3409	AAAGATTCAGAGGCTCTGAACTGTGGGAAAAACAGGAGTTTAGAAAAAAACCTGAA	3468
Db	3208	-----GAA	3210
QY	3469	ATTGCCCCAGGTTATTGCTTCTTACGCTGTCTATGTGTCCCGAACAACTCACCCCTGGCTCT	3528
Db	3211	ATTGCCCCAGGTTATTGCTTCTTATACCGCCACCGGCCCCGAGAGCTTACTCTTGCCCCCT	3270

OY	3529	GGGACGCTGATCTCTATCCGAGAAAAAGAACCCAGGATGATGGTGGAAAGAGAACTGCA	3588
Db	3271	GGTCAGCTGATTTTATTCGAAAAAGAACCCAGGATGATGGTGGAAAGAGAGCTGCA	3330
OY	3589	GCTCGAGGAAAAAAGCCGACAGTAAGGATGTTTCCAGCAAAATTAATGTCAAACTTTAGC	3648
Db	3331	GCACGTGGAAAAAAGCCGACAGTATGAGCTGGTTCCAGCTAATTAATGTAAAGTTCTTAGC	3390
OY	3649	CCCCGAACAAGAAAAATACCCCACTAGCTACCCAGACCCGACGTGACGACACTG	3708
Db	3391	CTTGGAGCAGAGAAAAATACCTCAACAGAGCCACTAAGTCAACAGCAATTAGCCGACTG	3450
OY	3709	TGCCAGTGTATCGGATGTACATTAACCCGCCCAAGACGATGACGAACTAGCTTACG	3768
Db	3451	TGCCAGTGTATTGGATGTACACTACACCCGCCGAAATGACGATGACTGGCTTCAAC	3510
OY	3769	AAAGCCAGATATCAACGCTCTCAACAAGAGAACCCGACTGCTGAAAAGAGAGTC	3828
Db	3511	AAAGGCCAGATATCAACGCTCTCAACAAGAGAACCCGACTGCTGAAAAGAGAGTC	3570
OY	3829	AGTGGGCAAGTTGGGCTCTTCCATCCCAATTAATGAAGAGTGAACAACAATGACCC	3888
Db	3571	AATGGAACAATGGGGCTCTTCCATCCCAATTAATGAAGAGTGAACAACAATGACCC	3630
OY	3889	AGCCGACATGATCATATGTTGTCATCCCCCTCAGGCTGAAAGTCTCAAGAGA	3948
Db	3631	AGCCGACATGATCATATGTTGTCATCCCCCTCAGGCTGAAAGTCTCAAGAGA	3690
OY	3949	CCCATATCCCATATACCTGCCCAAGAGATATGAGATATGACACTTGAATCTGAC	4008
Db	3691	CCCATATCCCATATACCTGCCCAAGAGATATGAGATATGACACTTGAATCTGAC	3750
OY	4009	TTGCAGATGATCACTACTGCGCTTCTAGTAGAAGAACTCACTCAGACAGTTACT	4068
Db	3751	TTCCAGATGATCACTACTGCGCTTCTAGTAGAAGAACTCACTCAGACAGTTACT	3810
OY	4069	CATTTGACCTTATGTTGCATGTGATGAAATGTCTGATCATCTGCTGCAAGGACAGC	4128
Db	3811	CATTTGACCTTATGTTGCATGTGATGAAATGTCTGATCATCTGCTGCAAGGACAGC	3870
OY	4129	--AAATTCAGAACTGCACAGGGATGGTGGCTTTGGGGCTTCCATGACATCAAC	4188
Db	3871	AAAAATTCAGAAACACACAGGGATGGTGGCTTTGGGGCTTCCATGATCTCAAT	3930
OY	4187	TGAC-CGGCCCCGCTTCAACCGGCGCTTCAATAGTTTAAAGTTATTTTAAATG	4245
Db	3931	TGACTTTCGCCCACTTGCACAGGTGCTTCAATAGTTTAAATTTTAAATGATA	3990
OY	4246	TATTTTACCTTTTATATAAATCTCAATCAATTACTCTTGGCTATTTTGGTTTACA	4305
Db	3991	TATTTTACCTTTTATATAAACAAATTAATTAATTAATTAATTAATTTTGGTTTACA	4050
OY	4306	AAACACCCACTATCAAGAGAGCTGTCTGGCGACGATTAATAATGCTGTTCCGGCGATA	4365
Db	4051	AAACACCCACTATCAAGAGAG-----CTGATGTGCTATTTAAATTTGTTCCAAAGT	4105
OY	4366	CCGTTAATCTAGAGCTTGTCTGTACTTTTCCGTTTGTCCAGTGTCCCAACCACTTG-T	4424
Db	4106	CCATTAATCTGAGACTGTATGATATTTTTCATTTTGTCCAGTGTTCAACAAATTAATGCT	4155
OY	4425	GTAGTTTGGGGCTGTTGCC---TGCGGTAAAGCAACAAGAGATGGGTGATACCTGTTT	4480
Db	4166	GCAGTTTGGGGCTTTTCCCTTACCAATGAGATGCAAGAGAGTTCAATATCTCTGTTT	4225
OY	4481	GAAAAATGTATGTAGACTGAGCCCTGACTATGGAAGGGATTATGCTTGTCTGACATC	4540
Db	4226	AAAGACGTATGAAATGAGCCCAATTAA--GGAAGTATGTTGTGTTTGTGTGATTC	4288
OY	4541	ACGTGTACC-TGTCCGATGATGACATCTGTACCGAAGAAATAGT-CTTCTCATATGC	4598
Db	4285	ACGTGTACCTTGTAGACATGTATATATCTGTATCATTAAGAAATTAATGTTCTTTCATAGG	4344

QY	4599	TAAACCCACACCGGTGATACAGTGGTCTGCATC-TATGCAATTCATTTACGTT--GCACAG	4655
Db	4345	CAAAGCTATTAACCTTTGATACATGCTCTTAATCAATATTGCAATTTATTTATTTTGCAACAG	4404
QY	4656	TGACCTTGTATCCACCTGAGGAAGCAC-CCATGTTTCCGTTGGTCTCAATGATACCTAG	4714
Db	4405	TGACCTTGTATCCACCAATGAGAAAGCACTGTGTTTTTGTTCGGTCTCAATTTATCTGG	4464
QY	4715	TTGTGCCCGGTGTTTTGTTTTTATTTTCAATCTGGCATGCTTTCACACCAATTAACTAGTA	4774
Db	4465	TTGAGTTGGTGTGTTTTGTTGGGTTTTTAAATTTTGGCGTGTTCATGATTAATATCAGT	4524
QY	4775	AGAGGCCAACCGCCACGGCGGTTTACATCATCATGATACCAACCGC-----TTAGTCTCTG	4829
Db	4525	AGACAAAC-ACCACCTGAGGTGTGTTACGATCAACATATCCACAGTCTCTTTTATGCTCTG	4583
QY	4830	TTAGCGTAG-TTTATTTCCAGTTGCTTTTATGGA-----TATCTTGAACAGTAACT	4883
Db	4584	TTACATGAAGTTTATTTATTCAGTACTTTTCTATGGAATGACCTATTTTGAACAGTAACT	4643
QY	4884	TCTTGAACAGAAAGATGTATATGAGTCTCCCTGCATTAATTTCCAGTGTTTACATTT	4943
Db	4644	TCTTGAACAGAAAGATGTATATGAGTCTCCCTGCATTAATTT-CCATGTTTACATTT	4702
QY	4944	TTTAACTATGACTGTGGGGGTGCTACAGATTAATATGAAATGGCGCTCTGTCCTGTG	5003
Db	4703	TTTAACTATGAGCTGTGGAATTTTCTACAGATTAATATGAAATGAGCTCATGTCCTGTTG	4762
QY	5004	TGTGTAACTGTGCTGTAGCTGAAGCCGTGTG---CCTTAGATATTAGTTGGAAGTCG	5060
Db	4763	TGTGTT-AGATATCTGTAGCTGAAGCCCTGTGTTGCTTTTAAACACTAATTGGAGCTC	4821
QY	5061	GGAAGAGAAATT 5071	
Db	4822	TCAATTAATAAT 4832	
RESULT 8			
AAS84763			
ID	AAS84763	standard; cDNA; 7435 BP.	
AC	AAS84763;		
XX			
DT	13-FEB-2002	(first entry)	
XX			
DE	DNA encoding novel human diagnostic protein #20567.		
XX			
KM	Human; chromosome mapping; gene mapping; gene therapy; forensic;		
KW	food supplement; medical imaging; diagnostic; genetic disorder; ss.		
XX			
OS	Homo sapiens.		
XX			
PN	WO200175067-A2.		
XX			
PD	11-OCT-2001.		
XX			
PF	30-MAR-2001; 2001WO-US08631.		
XX			
PR	31-MAR-2000; 2000US-0540217.		
XX			
PR	23-AUG-2000; 2000US-0649167.		
XX			
PA	(HYSE-) HYSEQ INC.		
XX			
PI	Drmanac RT, Liu C, Tang YT;		
XX			
DR	WPI; 2001-639362/73.		
XX			
DR	P-PSDB; ABG20576.		
XX			
New isolated polynucleotide and encoded polypeptides, useful in			
PT diagnostics, forensics, gene mapping, identification of mutations			
PT responsible for genetic disorders or other traits and to assess			
PT biodiversity -			
XX			

Db 3911 CAATTATGTGAAGCTGACACAGACATGGACCCAGCCAGCAATGTTGACACTTAC 3970
Qy 3915 ATCCCCCTCAGGCTTGAAGTCTCTCAAGAGACCCCTATCCATATC 3964
Db 3971 ATCTTGTGATATGTTGACCCCACTGAAGAAAGCAAGATATCATC 4020

RESULT 9
AA163825
ID AA163825 standard; cDNA; 3466 BP.
XX
XX AA163825;
AC
XX
DT 22-OCT-2001 (first entry)
DE Human polynucleotide SEQ ID NO 33.
XX
XX Human; antiarthritic; antirheumatic; antiproliferative; vasotropic;
KW cerebroprotective; nootropic; neuroprotective; antibacterial; virucide;
KW fungicide; ophthalmological; cytostatic; immunosuppressive; nootropic;
KW neuroprotective; antiallergic; hepatotropic; antidiabetic;
KW antiinflammatory; antiulcer; vulnery; anticonvulsant; antibacterial;
KW antiparasitic; cardiant; gene therapy; cancer; immune disorder;
KW cardiovascular disorder; neurological disease; infection; human; ss.
XX
OS Homo sapiens.
XX
XX WO20015308-A2.
PD
XX 02-AUG-2001.
PF 17-JAN-2001; 2001WO-US01309.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225457.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.

PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234224.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241121.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0244674.
PR 08-NOV-2000; 2000US-0244675.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.

PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249219.
 PR 17-NOV-2000; 2000US-0249220.
 PR 17-NOV-2000; 2000US-0249221.
 PR 17-NOV-2000; 2000US-0249222.
 PR 17-NOV-2000; 2000US-0249223.
 PR 17-NOV-2000; 2000US-0249224.
 PR 17-NOV-2000; 2000US-0249225.
 PR 17-NOV-2000; 2000US-0249226.
 PR 17-NOV-2000; 2000US-0249227.
 PR 17-NOV-2000; 2000US-0249228.
 PR 17-NOV-2000; 2000US-0249229.
 PR 17-NOV-2000; 2000US-0249230.
 PR 17-NOV-2000; 2000US-0249231.
 PR 17-NOV-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251889.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 XX WPI, 2001-488781/53.
 DR P-PSDB; AAM43519.
 XX
 PT New isolated nucleic acids and polypeptides, useful for diagnosing,
 PT treating and/or preventing human diseases and disorders -
 XX
 PS Claim 1. SEQ ID NO 33; 664bp + Sequence Listing; English.

XX The invention relates to human polynucleotides (AA163803-AA164012) and
 CC the encoded proteins (AAM434497-AAM43660) useful for preventing, treating
 CC or ameliorating medical conditions e.g. by protein or gene therapy. The
 CC genes were isolated from a range of human tissues disclosed in the
 CC specification. The nucleic acids, proteins, antibodies and (ant)agonists
 CC are useful in the diagnosis, treatment and prevention of: (a) cancer,
 CC e.g. breast and ovarian cancer and other cancers of the adrenal gland,
 CC bone, bone marrow, breast, gastrointestinal tract, liver, lung, or
 CC urogenital; (b) immune disorders e.g. Addison's disease, allergies,
 CC autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus,
 CC Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
 CC and parasitic infections.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at: ftp.wipo.int/pub/published_pct_sequences.
 XX
 XX Sequence 3466 BP; 1056 A; 829 C; 861 G; 710 T; 10 other;

Query Match 43.1%; Score 2188.8; DB 22; Length 3466;
 Best Local Similarity 84.7%; Pred. No. 0;
 Matches 2525; Conservative 0; Mismatches 427; Indels 30; Gaps 5;

QY 7 GAGGAGAGTGGAGCGCGCGGAGGCGGAGCGAGCTTGGTCTCCGTAGTACGCGGCT 66
 Db 35 GAGAAAGTGGAGCGCGCGGAGGAGCGAGCGTGGTCTCCGTAGTACGCGGCT 94
 QY 67 CGCAAGGAGCATCCCGAGCGGCTCCGAGCGCGCGGAGGAGCGAGCGAGCGCGCG 126
 Db 95 CGGAGGAGGAGTCCCGAGCGGCTCCGAGCGCGCGGAGCGCGCGCGCGCG 143
 QY 127 GGGATGTTGGCGCGGCTCGGAGCTCGCGGTTCTTCGCGCGCGCTCGCGAGCTGA 185
 Db 144 GGGATGTTGGCGCGGCTCGGAGCTCGCGGTTCTTCGCGCGCGCTCGCGAGCTGA 203

QY 186 TTTGTGAGGAGGCGCGCGCGCGCGCAACCGCGGAGATGAGGCGTGCATCAGCAAGTGA 245
 Db 204 TTTGTGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 263
 QY 246 ACGTAATGAGCAATGAGCTCACTTTTCCGACACTTTTGGGTGAGCTGAGTGTGGG 305
 Db 264 AAGTAACAGAACATGAGCTCACTTTTCCGACACTTTTGGGTGAGCTGAGTGTGGG 323
 QY 306 CAATACTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 365
 Db 324 CAATACTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 383
 QY 366 GGGATTTTATGAGTGCATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 425
 Db 384 TGGATTTTATGAGTGCATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 443
 QY 426 TGTCTTAGACAAATATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 485
 Db 444 TGTCTTAGACAAATATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 503
 QY 486 GGAATTTTTCATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 545
 Db 504 GGAATTTTTCATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 563
 QY 546 CACACTTCCCGCTGCATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 605
 Db 564 TGCACCTTCCCGCTGCATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 623
 QY 606 TATAGAGGAGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 665
 Db 624 TATAGAGGAGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 683
 QY 666 CATTCAGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 725
 Db 684 CATTCAGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 743
 QY 726 TCCCTGAGCTAACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 785
 Db 744 CCCCCTGAGCTAACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 803
 QY 786 AGCCATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 845
 Db 804 AGCCATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 863
 QY 846 TAAATTAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 905
 Db 864 TAAATTAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 923
 QY 906 TGTGCTGAGCTAACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 965
 Db 924 TGTGCTGAGCTAACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 983
 QY 966 GAGTGAACCTTAAC-----AGTCCCGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1013
 Db 984 GAGTGAACCTTAAC-----AGTCCCGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1043
 QY 1014 AAGTTTACCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1073
 Db 1044 AAGTTTACCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1103
 QY 1074 AAAATCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1133
 Db 1104 AAAATCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1163
 QY 1134 TCAGCACTGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1193
 Db 1164 CCAACCACTGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1223
 QY 1194 CTGCGGAGTGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1253
 Db 1224 ATCTGGAGTGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1283
 QY 1254 GCCGTGTCAGAGGAGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1310

Db	1284	ACCACTTTAGAAAGTGAACAACAATTAGAAAAAATTACTGTGACCTTTGAAGA	1343
Oy	1311	TAAGAAGCGGAGAACTTCGACGAGCAGTGTGAGCTGGAAGACCGCCGCAAGCGCT	1370
Db	1344	TAAGAAGCGGAGAACTTTGAACGTGGCACTTGAACTGGAGAAACGAAGCAAGCTCT	1403
Oy	1371	CTTGGAGCAGCAGCCCAAAAGACGAGAGCGGTGGCTCACTGCGCGCCGAGCAGA	1438
Db	1404	CCTGGAAACAGCGCCCAAGCAAGACAGACGCTTGCCACAGCTGGACGGCGGAGCAGA	1463
Oy	1431	GAGAAAAAGCGGAGCGCCAGAGCAGAGAGGCGCAAGCGGCAAGCTGGAGAACGA	1490
Db	1464	GAGGAAGAGCCTGAGCGCCAGAGCAAGAGCGCAAAAGACAATTGGAACGAGACGA	1522
Oy	1491	GCTGAGAAAGCAGCGGAGCTTGACCGGCGAGAGAGAGAGAGAGAGAGATCGA	1550
Db	1524	ACTGGAAGAGCAGCGGAGGCTTAGAACGCCAGAGAGAGAGAGAGAGAGAAATTGA	1583
Oy	1551	GAGCGCGAGCGCCGCAAAACGGGAATCTGGAAGGCAAGCAACCTTGAATGGGAACGGA	1610
Db	1584	GAGCGGAAGGCTGCAAAACGGGAATCTGGAAGGCAAGCAACCTTGAATGGGAACGGA	1643
Oy	1611	CCGAGACAGGAATCTCTGAATCAGAGGAACAAGAGACAGAGAGGCGACCGTGTCTGAA	1670
Db	1644	TCGAAGGCAAGAACTCTAAATCAAGAAACAAAGAAACAAAGAGCAATGTTGACTGAA	1703
Oy	1671	GGAAGAGAGAAAGCTCTGAGATTGAGTTAGAACTCTGAATGACAAAAGACATCACT	1730
Db	1704	AGCAAGAAAAAGACTTTGGATTGGATTGAAAGCTTAAATGATTAAGAAAGATCACT	1765
Oy	1731	AGAAAGAAAATCTTCAGATATCAGGTGTGCACTGGCAACCCAGAGCAAGAAATTGAGAG	1790
Db	1764	AGAAAGGAAAATCTTCAGATATCAGATGTGCAATTGACACCCAAAGCAAGAAATTGAGAG	1823
Oy	1791	CACGGAACAGCTTAGAGGCTTAAGATTGTGAATTCACCACTTACAGCAGCAATTCGA	1855
Db	1824	CACAAACAAATCTTAAGAGATTGAGAAATTCGCCAAATTCACCACTTACAGCAACAATTCA	1883
Oy	1851	GGAATCTCAGCAAAATGCTTGGAGAACTTAATCCAGAGAAACAGATATCAGTACCACTT	1910
Db	1884	GGAATCTCAGCAAAATGCTTGGAGAACTTAATCCAGAGAAACAGATATCAGTACCACTT	1943
Oy	1911	AAAAAAGTCCAGCAAAACAGTTTGCATAGAGACTCGCTCTTCAACCTTCAAAAGACCTT	1970
Db	1944	AAAAAAGTCCAGCAAAACAGTTTGCATAGAGACTCACTTGTTCACCTTAAAGACCTT	2003
Oy	1971	GGAACCAAGAGCTGGCCCGGCACAGCTCCGGAGCAGCTGGAACAGGTGAGAGAGA	2030
Db	2004	AGAAACCAAGAACTAGCTCCGACACACTTAGAACCAACATCGATGAAAGTGGAGAAAAGA	2063
Oy	2031	GACCAAGTCAAAAGCTGCAAGAGATTGATGTTTCAACAACCAAGCTGAAGAACTAGAGA	2090
Db	2064	AACCTGATCAAAACTACAGGAATTTGATTTTCAATATTCAGCTGAAGAACTTAGAGA	2122
Oy	2091	GATACATAGCAAAACAGCACTTCAGAGCAGAGACTCCCTGAGGAGCGCGCATGAAGCA	2155
Db	2124	AATACACATAGCAAAACCACTTCAGAGGAAAGTTCATGAGGCTGAACGATGAAGCA	2188
Oy	2151	GAAAGAGCAGAGAGAGAGAGCTTGAAGTTAGAGAAAGCAAAAGAAAGCTCAGAGACG	2210
Db	2184	GAAAGAAACAGAGACCAAGAGATCATGAAATTGAAAAAACAAGAAAGAAAGCCCAAGACG	2243
Oy	2211	AGTTAGAGAAAGGGAACAAGCAATGGCTGGAGAGTTGACAGCAGAGG---GAGCAACCAAG	2266
Db	2244	AGCTCAGAGAAAGGGAACAAGCAATGGCTGGAGATTTGACAGCAGGAGACAGCAATCAAG	2303
Oy	2268	CCCCCGGAAACCCCAAGAGAGAGACGCTGAAGAGGGAACAAGTCTCAGAGAAAGAGA	2327
Db	2304	ACCAAGAAAACTCCAGAGAGAGAAAACTGAAAAAGGAGAGAGATGTCTAAAAAAGAGAG	2365
Oy	2328	GGCGGAAGAGAGCCAAAGCCGGAATGCAAGACAAGCAAGTCCGCTTTTCATCCCA	2387

Db	2364	TGGCAGAGAAAAAGCCAAACAGGAAGCAACAGACAGCTGGGTGGCTTTTCATCAACA	2425
Qy	2388	TCAGGAGCCAGCTAAAGTCTGGCCACCAGGACCCCTGTCTACCAACAGAAAGGCCGCT	2447
Db	2424	CCAAAGAACACAGCTAAGCCAGCTGTCCAGGACCCTGGGTCCACTCCAGAAAAAGTCCACT	2483
Qy	2448	TACCACTTCTGCAACAGGAGAGTGTAAAAAGTGTATATTACCGAGCGCTGTACCCCTTTGA	2507
Db	2484	TACCACTTCTGCAACGAGAAAAATGTAAAAAGTGTGTATTACCGGACACTGTACCCCTTTGA	2543
Qy	2508	ATCCGAAAGTACGATGATGATCACCATCTCAGCCAGAGAGATPATGTATGTGTGATGAAG	2567
Db	2544	ATCCGAAAGCCATGATGAATATCATCTATCCAGCAGAGAGACATATGATGTGTGATGAAG	2603
Qy	2568	CCAGACTGAGAGCCAGGATGAGCTTTGAGAGAGCTGAAGGGGAAGCGGATGTTCC	2627
Db	2604	CCAAACTGAGAGAACCCGGCTGAGCTTTGAGAGAGATTAAAGAAAGACAGGGTGGTCCC	2663
Qy	2628	TGCAAACTATGCAAAAAAGATTTCCAGAAATAGAGTTTCCACTCCAGCCAAACCAATGAC	2687
Db	2664	TGCAAACTATGCAAGAAATATCCAGAAATAGAGTTTCCCTCCATCAGTGAACCAATGAC	2723
Qy	2688	CGATCTGACATCTGCGCCCTGCCCCCAAACTGAGCTCTGCTGTGAGAGCCCTGCTCTTGGC	2747
Db	2724	TGATTCAAACATCTGCTCCCTGCCCCCAACTGAGCTCTGCTGTGAGAGCCCTGCTCTTGGC	2760
Qy	2748	AGTGAACCTCTTCTGAGCCCTCCACAAACCCCAACAACTGGGACAGATTCACTTCCACGTG	2807
Db	2784	AGTAACCTCTTCTGAGAGCCCTCCACAGACCCTTAATATCTGGGCGCAGCTTCACTCCAGTG	2843
Qy	2808	GCCCAAGCTCTCAAAACGAGAAACCGAACAACCTGGATACGTGGGTGGCTCAGCC	2867
Db	2844	GCCCAAGCTCTCAAAACGAGAAACCGAACAACCGATTAACCTGGATACATGGGACGCCAGCC	2903
Qy	2868	TTCTGTGACCGATACCTTAGTGTGGCGACATTCAGGAGAGATCAGCTTTACCCAGACAC	2927
Db	2904	CTCTTTCACCGTTCCAAATGTGCGGCGACGTTAAGGACAGAGTCCCGCTTTACTCCAGCCAC	2963
Qy	2928	AGCCACTGGCTCTTCCCATCTCCCGTCTCTGGGACAGGGTGA	2969
Db	2964	GGCCACTGGCTCTTCCCGTCTCTGTGCTGTAGGCGCAGGCTTA	3005
RESULT 10			
ABK43498			
ID	ABK43498	standard; cDNA; 319 BP.	
AC	ABK43498;		
XX			
DT	05-JUN-2002	(First entry)	
XX			
DE	DNA encoding novel central nervous system protein #78.		
XX			
Kw	Central nervous system; CNS; autoimmune disease; rheumatoid arthritis;		
Kw	hyperproliferative disorder; neoplasm; cardiovascular disorder;		
Kw	cardiac arrest; cerebrovascular disorder; ischemia; angioneuroma;		
Kw	nervous system disorder; Alzheimer's disease; AIDS; ocular disorder;		
Kw	acquired immunodeficiency virus; dysphagia; gastrointestinal disorder;		
Kw	adenocarcinoma; reproductive system disorder; testicular feminization;		
Kw	endocrine disorder; diabetes; cancer; leukemia; neovascularization;		
Kw	respiratory disorder; renal disorder; kidney failure; blood disorder;		
Kw	myocardial infarction; wound healing; cell proliferation; skin aging;		
Kw	food additive; food preservative; gene therapy; gene; ss.		
XX			
OS	Homo sapiens.		
XX			
PN	WO20015318-A2.		
XX			
PD	02-AUG-2001.		
XX			
PF	17-JAN-2001; 2001WO-US01332.		
XX			
PR	31-JAN-2000; 2000US-0179065.		

```
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0186974.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214866.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231744.
PR 08-SEP-2000; 2000US-0231744.
PR 08-SEP-2000; 2000US-0231744.
PR 08-SEP-2000; 2000US-0231968.
PR 12-SEP-2000; 2000US-0232081.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234423.
PR 21-SEP-2000; 2000US-0234424.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235634.
PR 27-SEP-2000; 2000US-0235636.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236602.
PR 02-OCT-2000; 2000US-0237037.

PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-02559678.

(HUMA-) HUMAN GENOME SCI INC.
PI
XX
PI Rosen CA, Barash SC, Ruben SW,
XX
XX MPI: 2001-581633/65.
DR P-PSDB; AA087168.
XX
XX New isolated nucleic acid encoding a protein for diagnosing,
PT preventing, treating or ameliorating medical conditions and used as
PT food additives or preservatives -
```



```

QY 1835 TACAGAGCAGTGTGAGGAATCTCAGCAATGCTTGAAGACTTATTCAGAGAAACAGA 1894
    |||||
Db 1750 TACAGAACATTTACAGGAATCTCAGCAATGCTTGAAGACTTATTCAGAGAAACAGA 1809
QY 1895 TACTCAGTGAACGATTAAACAAAGTCCAGAGAAACAGTTTGCATAGAGACTCGCTTCTTA 1954
    |||||
Db 1810 TACTCAGTGAACGATTAAACAAAGTCCAGAGAAACAGTTTGCATAGAGACTCGCTTCTTA 1869
QY 1955 CCCTCAAAAGAGGCTTGAAGCAAAAGAGTGGCCCGGACAGCTCCGGGAGCAGCTGG 2014
    |||||
Db 1870 CACTTAAAGAGGCTTGAAGCAAAAGAGTGGCCCGGACAGCTCCGGGAGCAGCTGG 1929
QY 2015 ACCAGGTGAGAGAGAGACAGGTCAGCAAGTGCAGAGAGATTGTTTCAACAAACAGC 2074
    |||||
Db 1930 ATGAAGTGAAGAAAGAAACCTAGATCAAACTACAGAGATGATATTTTCAATTAATCAGC 1989
QY 2075 TGAAGGAATGAGAGAGATACATAGCAAAACAGCAACTCCAGAAAGCAGAGGTCCTGAGAG 2134
    |||||
Db 1990 TGAAGGAATGAGAGAAATACATATAGCAAACTCCAGAAAGCAAAAGTCCATGAGAG 2049
QY 2135 CAGCGGAGCTGAAGCAAGAAAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2194
    |||||
Db 2050 CTGAAGGAGCTGAAGCAAGAAAGAAACAGAAAGAGATCATAGAAATTAAGAAACAAAG 2109
QY 2195 AAGACGCTGAGAGAGAGAGTTCAGAGAAAGGAGCAAGCAATGGCTGAGAGATGTCAGAGAG 2254
    |||||
Db 2110 AAGAAACCCAAAGACAGAGCTCAGAGAAAGGAGCAAGAGAGAGAGAGAGAGAGAGAG 2169
QY 2255 AG--GAGCAGCAGCGCCCGGAAACCCACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2311
    |||||
Db 2170 AGGACGAGCATCAGAGAGCCAAAGAACTCCACAGAGAGAGAGAGAGAGAGAGAGAGAG 2229
QY 2312 GTGTCAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2371
    |||||
Db 2230 GTGTCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2289
QY 2372 GGGTTTCCATCCGAGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2431
    |||||
Db 2290 GGGTTTCCATCCATCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2349
QY 2432 CAGAGAAAGAGCGGCTTACATTTCTGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2491
    |||||
Db 2350 CAGAAAGAGGTCCACTTACATTTCTGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2409
QY 2492 CGGTGATCCCTTTGATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2551
    |||||
Db 2410 CACTGATCCCTTTGATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2469
QY 2552 TCATGTGATGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2611
    |||||
Db 2470 TCATGTGATGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2529
QY 2612 AGACGGAGATGATTCCTGCAAACTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2671
    |||||
Db 2530 AGACGGAGATGATTCCTGCAAACTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2589
QY 2672 CAGCCAAACAGTGAACGATCTGATCTGCCCCCTGCCCCCAAACTGGCTCTGCTGAGAG 2731
    |||||
Db 2590 CAGTGAACAGTGAACGATCTGATCTGCCCCCTGCCCCCAAACTGGCTCTGCTGAGAG 2649
QY 2732 CCCGCTCTCTTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2791
    |||||
Db 2650 CCCGCTCTCTTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2709
QY 2792 ACTTCAGTTTCAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2851
    |||||
Db 2710 ACTTCAGTTTCAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2769
QY 2852 CGTGGGAGGCTCAGCTTCTGACCGATCTTATGTCGAGAGAGAGAGAGAGAGAGAGAGAG 2911
    |||||
Db 2770 CAGTGGGAGGCTCAGCTTCTGACCGATCTTATGTCGAGAGAGAGAGAGAGAGAGAGAGAG 2829

```

```

QY 2912 CCTTACCCAGAGCAGAGCAGGCACTGAGCTCTCCCATCTCCCGTCTGGGCAAGGTGA 2969
    |||||
Db 2830 CCTTACTCAGAGCAGGCACTGAGCTCTCCCGTCTCTGTGAGGCAAGCTTGA 2887
    |||||

RESULT 11
AAK94139
ID AAK94139 standard; cDNA; 2131 BP.
XX
AC AAK94139;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human full-length cDNA, SEQ ID NO: 2646.
XX
KM Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
XX
OS Homo sapiens.
XX
PN EP130094-A2.
XX
PD 05-SEP-2001.
XX
PF 07-JUL-2000; 200BP-0114089.
XX
PR 08-JUL-1999; 993P-0194486.
PR 11-JUN-2000; 2000P-0118774.
PR 02-MAY-2000; 2000P-0183765.
XX
PA (HELI-) HELIX RES INST.
PI Ota T, Nishikawa T, Ieogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
DR WPI; 2001-524255/58.
DR P-PSDB; AAM93229.
XX
PT 830 Primers useful for synthesizing full length cDNA clones and their
PT use in genetic manipulation -
XX
PS Claim 8; SEQ ID NO 2646; 1380bp + sequence listing; English.
XX
CC The invention relates to primers for synthesizing full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been
CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
CC molecules have been determined. Primers for synthesizing the full length
CC cDNA are useful for clarifying the function of the protein encoded by
CC the cDNA. The full length clones were obtained by construction of full
CC length enriched cDNA libraries that were synthesised by the oligo-capping
CC method. The primers enable the production of the full length cDNA easily
CC without any special methods. The present sequence is a full length
CC human cDNA of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in CD-ROM format directly from EPO.
XX
SQ Sequence 2131 BP; 694 A; 469 C; 526 G; 442 T; 0 other;

Query Match 31.7%; Score 1610; DB 22; Length 2131;
Best Local Similarity 85.5%; Pred. No. 0;
Matches 1832; Conservative 0; Mismatches 295; Indels 15; Gaps 3;

QY 45 GTTGTCTCCGTATGAGGCGGCTTCGCAAGGAGACATCCGAGCGGCTCCGGAGCGCCGG 104
    |||||
Db 1 GTTGTCTCCGTATGAGGCGGCTTCGCAAGGAGAAATCCGAGCGGCTCCGGAGCGGA--- 57
    |||||

QY 105 GAGGACGAGCGGCGGCGGCGGAGATGATGCGCGGCTGCGGAGCTGCGGCTGCTCTGCGC 164
    |||||
Db 58 -----CAGAGAGCGGCGGCGGAGATGATGCGCGGCTGCGGCTGCTCTGCGCCTCTCC 109
    |||||

QY 165 -CGGCGTGGCGGCTGCACTGATTTGTGAGAGGAGCGGCGGCGGCAAGCGGCGGAGAT 223
    |||||
Db 110 AGCGGCGGCTGAGCGGCACTGATTTGTCTCCCTGGGCGGAGGCGGAGCCCGCGGAGAT 169
    |||||

```

224 GAGGCGTGCATCAGCAGAGTGAAACGTAATAGAACCATGGCTCAGTTTCCACACCTTTG 283
 170 GAGGCGTGCATCAGCAGAGTGAAACGTAATAGAACCATGGCTCAGTTTCCACACCTTTG 229
 284 GTGGTAGCCTGAGATCTTGGGCGCATTAAGTGAGGAAAGGGCCAGAGTACACAGAGT 343
 230 GTGGAGCCTGGATATCTGGGCGCATTAAGTGAGGAAAGGGCCAGAGTACACAGAGT 289
 344 TCCCTTAGCCTGAGAGCGGATAGGCGGATTTATTAAGTGAGTACAGCGAGAACTTTT 403
 290 TCCATAGTTTAAAGCCAAATATCTGATTCATTAAGTGAGTACAGCGGAACTTTT 349
 404 TCCATCTGGGTACCTCAGCTGTCTTAGCACAATATAGGCGGTACGCGACATGATA 463
 350 TTCAATCTGGGTACCTCAGCTGTCTTAGCACAATATAGGCGGTACGCGACATGATA 409
 464 AGCATGAAAGATGATCAAGTGAAATTTCCATAGCCATGAGCTTATCAAACTGAAGC 523
 410 ATGATGAAAGATGATCAAGTGAGATTTCCATAGCTATGAACTTATCAAACTGAAGC 469
 524 TACAAAGATATCAAGTCTCCCTCCACACTTCCCTGTATGAAAGACCAAGTGGCTA 583
 470 TACAAAGATATCAAGTCTCCCTCCACACTTCCCTGTATGAAAGACCAAGTGGCTA 529
 584 TTTCCAGTGCACACCATTTGGTATAGAGGAGATTGTAGCATGCGACCACTACAGCTG 643
 530 TTTCTAGGCGACACCATTTGGTATAGAGGAGATTGTAGCATGCGACCACTACAGCTG 589
 644 TTTCTAGGCGACACCATTTGGTATAGAGGAGATTGTAGCATGCGACCACTACAGCTG 703
 590 TTTCTAGGCGACACCATTTGGTATAGAGGAGATTGTAGCATGCGACCACTACAGCTG 649
 704 TTTCTAGGCGACACCATTTGGTATAGAGGAGATTGTAGCATGCGACCACTACAGCTG 763
 650 TTTCTAGGCGACACCATTTGGTATAGAGGAGATTGTAGCATGCGACCACTACAGCTG 709
 764 TTTCTAGGCGACACCATTTGGTATAGAGGAGATTGTAGCATGCGACCACTACAGCTG 823
 710 TTTCTAGGCGACACCATTTGGTATAGAGGAGATTGTAGCATGCGACCACTACAGCTG 769
 824 TTTCTAGGCGACACCATTTGGTATAGAGGAGATTGTAGCATGCGACCACTACAGCTG 883
 770 TTTCTAGGCGACACCATTTGGTATAGAGGAGATTGTAGCATGCGACCACTACAGCTG 829
 884 TTTCTAGGCGACACCATTTGGTATAGAGGAGATTGTAGCATGCGACCACTACAGCTG 943
 830 TTTCTAGGCGACACCATTTGGTATAGAGGAGATTGTAGCATGCGACCACTACAGCTG 889
 944 TTTCTAGGCGACACCATTTGGTATAGAGGAGATTGTAGCATGCGACCACTACAGCTG 1003
 890 TTTCTAGGCGACACCATTTGGTATAGAGGAGATTGTAGCATGCGACCACTACAGCTG 949
 1004 TTTCTAGGCGACACCATTTGGTATAGAGGAGATTGTAGCATGCGACCACTACAGCTG 1063
 950 TTTCTAGGCGACACCATTTGGTATAGAGGAGATTGTAGCATGCGACCACTACAGCTG 1009
 1064 TTTCTAGGCGACACCATTTGGTATAGAGGAGATTGTAGCATGCGACCACTACAGCTG 1123
 1010 TTTCTAGGCGACACCATTTGGTATAGAGGAGATTGTAGCATGCGACCACTACAGCTG 1069
 1124 TTTCTAGGCGACACCATTTGGTATAGAGGAGATTGTAGCATGCGACCACTACAGCTG 1183
 1070 TTTCTAGGCGACACCATTTGGTATAGAGGAGATTGTAGCATGCGACCACTACAGCTG 1129
 1184 TTTCTAGGCGACACCATTTGGTATAGAGGAGATTGTAGCATGCGACCACTACAGCTG 1243
 1130 TTTCTAGGCGACACCATTTGGTATAGAGGAGATTGTAGCATGCGACCACTACAGCTG 1189
 1244 TTTCTAGGCGACACCATTTGGTATAGAGGAGATTGTAGCATGCGACCACTACAGCTG 1300
 1190 TTTCTAGGCGACACCATTTGGTATAGAGGAGATTGTAGCATGCGACCACTACAGCTG 1249
 1301 TTTCTAGGCGACACCATTTGGTATAGAGGAGATTGTAGCATGCGACCACTACAGCTG 1360

1250 CGTTTGAAGATTAAGAGCGGAGAACTTTGAACCTGGCAACTGGAGAAACGAA 1309
 1361 GCCAAGCGCTCTTGAAGAGAGAGCGCAAAAGAGAGAGCGGTTGCTTCAAGTGAAGCGG 1420
 1310 GGCAGGCTCTCTGGAACAGAGCGCAAGAGAGAGAGCGGCTGGCCAGCTGGAAGCGG 1369
 1421 CCGAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1480
 1370 CCGAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1429
 1481 TGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1540
 1430 TGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1489
 1541 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1600
 1490 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1549
 1601 GCGAAGCGAAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1660
 1550 GCGAAGCGAAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1609
 1661 TGGTCTGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1720
 1610 TGGTCTGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1669
 1721 AGCATCACTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1780
 1670 AGCATCACTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1729
 1781 AAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1840
 1730 AAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1789
 1841 AGCATCACTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1900
 1790 AGCATCACTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1849
 1901 GTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1960
 1850 ATGACCAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1909
 1961 AAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2020
 1910 AAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1969
 2021 TGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2080
 1970 TGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2029
 2081 AACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2140
 2030 AACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2089
 2141 GACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2182
 2090 GACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2131

RESULT 12
 AA234573
 ID AA234573 standard; cDNA; 2079 BP.
 XX AA234573;
 AC
 XX
 DT 01-FEB-2000 (first entry)
 XX
 DE Human SH3D1A cDNA clone 5.
 XX
 KW SH3D1A gene; human; Down's syndrome; leukemia; cancer;
 KW megakaryocytic abnormality; myeloproliferative disorder;
 KW platelet disorder; neutral disorder; thrombocytopenia;


```

QY 1381 CAGCGCAAGAGCAGAGCGGTTGGCTCAGCTGAGCGCCCGCCGACAGAGAGAAAGAG 1440
DB 1261 CAGCGCAAGAGCAGAGCGGTTGGCTCAGCTGAGCGCGCGAGAGAGAGAGAGAG 1320
QY 1441 CCGGAGCGCCAGCAGAGAGAGCGGCGAGCTGAGCTGAGAGAGCAGCTGAGAGAG 1500
DB 1321 CCGTGAAGCGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
QY 1501 CAGCGGAGCTGAGAGCGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560
DB 1381 CAGCGGAGCTGAGAGCGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
QY 1561 GCGCGCAAGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620
DB 1441 GCTGCAAGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500
QY 1621 GAACTCTGAGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1680
DB 1501 GAACTCTGAGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560
QY 1681 AAGACTCTGAGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1740
DB 1561 AAGACTCTGAGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620
QY 1741 CTTGAGAGTATCAGAGTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1800
DB 1621 CTTGAGAGTATCAGAGTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1680
QY 1801 TCTGAGAGTATCAGAGTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860
DB 1681 TCTGAGAGTATCAGAGTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1740
QY 1861 CAAATCTTGAAGAGCTTATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1920
DB 1741 CAAATCTTGAAGAGCTTATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1800
QY 1921 CAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1980
DB 1801 CAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860
QY 1981 GAGCTGCGCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2040
DB 1861 GAGCTGCGCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1920
QY 2041 AAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2100
DB 1921 AAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1980
QY 2101 AAACAGCACTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2160
DB 1981 AAGCAACAACCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2040
QY 2161 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2198
DB 2041 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2078

```

```

XX 07-FEB-2001.
PD 28-JUL-2000; 2000EP-0116126.
XX 29-JUL-1999; 99JP-0248036.
XX 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX (HELI-) HELIX RES INST.
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.
XX Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX Claim 8; SEQ ID 15658; 2537pp + CD ROM; English.
XX The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
SQ Sequence 2131 BP; 691 A; 448 C; 572 G; 420 T; 0 other;
Query Match 27.6%; Score 1404.2; DB 22; Length 2131;
Best Local Similarity 81.0%; Pred. No. 0;
Matches 1740; Conservative 0; Mismatches 283; Indels 126; Gaps 4;
QY 7 GAGGAGAGTGTGAGCGGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 66
DB 105 GAGAGAGAGTGTGAGCGGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 164
QY 67 CGCAAGGAGAGATCCGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 126
DB 165 CGCAGAGAGAGATCCGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 213
QY 127 GGGATGTGTGCGGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 185
DB 214 GGGATGTGTGCGGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 273
QY 186 TTTGTGTGAGGAGGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 245
DB 274 TTTGTGTGAGGAGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 333
QY 246 ACGTAATAGAACATGAGCTGAGTTCCACACCTTTGAGTGTGAGCTGAGTGTCTGGGC 305
DB 334 AAGTAACAGAACATGAGCTGAGTTCCACACCTTTGAGTGTGAGCTGAGTGTCTGGGC 393

```

```

RESULT 13
AAH16578
ID AAH16578 standard; cDNA; 2131 BP.
XX AAH16578;
XX 26-JUN-2001 (first entry)
XX Human cDNA sequence SEQ ID NO:15658.
XX Human; primer: detection; diagnosis; antisense therapy; gene therapy; ss.
XX Homo sapiens.
XX EP1074617-A2.

```


FT CDS 2..1927
 FT /tag= a
 XX MO9953062-A2.
 XX 21-OCT-1999.
 XX 16-APR-1999; 99MO-US08371.
 XX 16-APR-1998; 98US-0082007.
 XX (CEDA-) CEDARS SINAI HEALTH SYSTEM.
 XX Korenberg JR, Chen X;
 XX WPI, 1999-633829/54.
 XX P-PSDB; AAY32158.
 PT Nucleic acid from the human SH3D1A gene and its products, useful for
 PT the diagnosis and treatment of myeloproliferative disorders and
 PT leukaemia -
 PS Claim 2; Fig 14; 99p; English.
 CC This is the nucleotide sequence of a non-full-length cDNA (clone
 CC 9) corresponding to a novel human SH3 gene, termed the SH3D1A
 CC gene, that contributes to the development of platelets and the
 CC pathogenesis of leukaemias, both in general and in particular those
 CC involving the megakaryocytic lineage. The SH3D1A gene maps to the
 CC small candidate region for low platelets on chromosome 21.
 CC Sequencing of 5 different sizes of cDNA clone from foetal brain
 CC (see A4234570-74) suggests that at least 3 isoforms exist. The
 CC invention provides methods for the diagnosis and treatment of
 CC megakaryocytic abnormality, myeloproliferative disorder, platelet
 CC disorder, acute leukaemia, neural disorders, thrombocytopenia,
 CC platelet disorder on chromosome 21, low platelets in deletion for
 CC 21, association of gains in chromosome 21 with leukaemias, neural
 CC abnormalities, dysfunctions and disorders including brain
 CC malformations and corresponding cognitive dysfunctions,
 CC microcephaly, lissencephaly, and colpocephaly. Methods are also
 CC provided for: suppressing cells unable to regulate themselves;
 CC screening for a somatic alteration in the SH3D1A gene; monitoring
 CC the progress and adequacy of a treatment; monitoring tumour risk
 CC progress or megakaryocytic abnormality; myeloproliferative disorder,
 CC haematopoietic disorder, platelet disorder or leukaemia; and
 CC treatment of a subject (including a prenatal subject) having
 CC megakaryocytic abnormality, myeloproliferative disorder,
 CC platelet disorder, leukaemia or neural disorder using a
 CC nucleic acid that expresses SH3D1A or its antisense nucleic acid.
 XX
 SQ Sequence 3231 BP; 1004 A; 721 C; 712 G; 794 T; 0 other;
 Query Match 27.4%; Score 1395.4; DB 20; Length 3231;
 Best Local Similarity 70.7%; Pred. No. 0;
 Matches 2359; Conservative 0; Mismatches 506; Indels 470; Gaps 19;

DB 241 AGACCACTGGAATGAGGAGAAAGAACTAGATCAAAAATACAGAGATTGATATTT 300
 QY 2064 CAACACCAGCTGAAGGAACTGAGAGATTCATGCAACACGAACTCCAGAAAGAG 2123
 DB 301 CAATATACAGCTGAAGGAACTGAAGAAATACAAATATAGCAACCTCCAGAGCAAA 360
 QY 2124 GTCCCTGAGGACCGGAGCTGAAGCGAAAGACAGAGAGAGAGGCGCTGAGTTAGA 2183
 DB 361 GTCCATGAGGCTGAACGACTGAACAGAAAGAAACAGAAAGCAATCATGAAATTAGA 420
 QY 2184 GAACAAAGAGAGACGCTCAGAGACGATTCAGAGAAAGGACAAAGCATGCTGAGCA 2243
 DB 421 AAAACAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
 QY 2244 TGTGACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2300
 DB 481 TGTGACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540
 QY 2301 GAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2360
 DB 541 AAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
 QY 2361 CAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2420
 DB 601 CAAGCTGGGTGGCTTTTCCATCAACACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
 QY 2421 CTGCTTCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2480
 DB 661 CTGGCTCACTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
 QY 2481 AATATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2540
 DB 721 GTATTAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
 QY 2541 AGAGATATAGTCAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2600
 DB 781 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840
 QY 2601 GCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2660
 DB 841 ATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
 QY 2661 GGTTCACAGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2720
 DB 901 GGTTCACAGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960
 QY 2721 TCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2780
 DB 961 CTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020
 QY 2781 CAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2840
 DB 1021 TAACTGGGCGCACTTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 1080
 QY 2841 CAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2900
 DB 1081 TAACTGGGCGCACTTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 1140
 QY 2901 GCAAGATCAAGCTTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 2960
 DB 1141 GCAAGATCAAGCTTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 1200
 QY 2961 CCAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3020
 DB 1201 CCAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
 QY 3021 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3080
 DB 1261 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
 QY 3081 GTGTGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3140

Db 1321 GTGGTGGTTGGAAGTTCAAGGTCAGAAAGGTTGGTTCGCCAAGTCTTACGTGAACCT 1380
QY 3141 CATTTCAGGGCCCGTAAGAAATCCACAGCATCGATCTGCGCCCTACCTGAAGTCTGCG 3200
Db 1381 CATTTCAGGGCCCGTAAGAAATCCACAGCATCGATCTGCGCCCTACCTGAAGTCTGCG 1440
QY 3201 TAGCTTAAAGAGATGGCTTCCCGGCGCCCAAGCCAGCATTCGCGAGAAAGATTAT 3260
Db 1441 TAGCTTAAAGAGATGGCTTCCCGGCGCCCAAGCCAGCATTCGCGAGAAAGATTAT 1492
QY 3261 TGCCATGTACACATACGAGAGTTCTGAGCAAGAGATTTAACCTTTCAGCAGGGAGAT 3320
Db 1493 ----- 1492
QY 3321 GATTGTGTTACCAAGAAAGATGTGACTGTGTGACGGGACGAGACGGTGGCGCAAGTCCGG 3380
Db 1493 ----- 1492
QY 3381 AGTCTTCCCTTCTAAGTATGTAGAGCTTAAAGATTCAGAGGCTCTGAACTGTGGGAA 3440
Db 1493 ----- 1492
QY 3441 AACAGGAGTTTAGAAGAAAAACCTGAATTTGCCCAAGTTATTGCTTCTACCGCTGTAC 3500
Db 1493 -----GAAATTTGCCCAAGTTATTGCTTCTACACCGCCAC 1527
QY 3501 TGGTCCCGAACATCACTGCTGCTCTGCGGAGCTGATCTGATCCGAAAAAGAACCC 3560
Db 1528 CGGCCCCGAGACCTCACTGCGCCCTGCTGAGCTGATTTTGTATCCGAAAAAGAACCC 1587
QY 3561 AGGTGATGGTGGAGAGAGAACTGCAAGCTCGAGGAAAAAGCCAGATAGAGTGT 3620
Db 1588 AGGTGATGGTGGAGAGAGAGCTGCAAGCACTGGAAAAAGCCAGATAGAGTGT 1647
QY 3621 TCCAGCAATTTATGTCAAACTTCTAAGCCCCGGAACAAGCAAAATCACTCAAGT 3680
Db 1648 CCAGCTAATTTATGTCAAACTTCTAAGCCCCGGAACAAGCAAAATCACTCAAGT 1707
QY 3681 ACCCAAGCCGAGTGCAGGACGACGATGTGCGCAGGTGATGGATGTACATTAACCCG 3740
Db 1708 ACTTAAGTCAACAGCATTTAGCCGAGTGTGCGCAGGTGATGGATGTACATTAACCCG 1767
QY 3741 CCAGACAGTACGAACTAGCTTTCAGCAAAAGCCAGATCATCAAGTCTCTCAACAGGA 3800
Db 1768 GCAAGATGACATGAGCTGCTTCAACAGGCGCAGATCATCAAGTCTCTCAACAGGA 1827
QY 3801 GGAACCGGACTGTGTGAAAAGAGAGTCACTGTGGCAAGTTGGCTCTTCCCATCAATTA 3860
Db 1828 GGAACCGGACTGTGTGAAAAGAGAGTCACTGTGGCAAGTTGGCTCTTCCCATCAATTA 1887
QY 3861 TGTAAAGCTGACCAACAGACATGTGACCCAGCCAGCAATGATATATGTTTCCATCCC 3920
Db 1888 TGTAAAGCTGACCAACAGACATGTGACCCAGCCAGCAATGATATATGTTTCCATCCC 1947
QY 3921 CCTTCAGGCTTGAAGTCTCTCAAGAGACCCACTATCCATCATCTGCGCCAGAGGATG 3980
Db 1948 CCTTCAGGCTTGAAGTCTCTCAAGAGACCCACTATCCATCATCTGCGCCAGAGGATG 1962
QY 3981 ATGGAGATGACACTTGTATCATGTGACTTTCAGAGATGATCACTGCTTCTGAGTA 4040
Db 1963 ----- 1962
QY 4041 GAAAGAACTACTGCAAGACAGTTTACCTCACTTGGACCTTAGTGTGATGGAATGT 4100
Db 1963 ----- 1962
QY 4101 CTGAGTCACTGCGTGCAGAGGCAAGACAAATTCAGAACTGCAAGGATGATGATCC 4160
Db 1963 -----GTTCT 1967
QY 4161 TTTGGGCTTTTCTAGTCACTGACAGTAC-CGGCCCGGCTTCAACAGGCGCTTTCAA 4219
Db 1968 TTTGGGCTTTTCTAGTCACTGACAAATTCATTTCCCGCACTTTTGACAGGTCCTTCA 2027

QY 4220 TAGTTTAAGATATTTTAAATNGTATTTTACCTTTTAAATTAATCTCAATCAAT 4279
Db 2028 TAGTTTAAGATATTTTAAATNGTATTTTACCTTTTAAATTAATCTCAATCAAT 2087
QY 4280 ACTTCTTGCTATTTTGGTTTAAACAAACACCCATATCAAGAGTGCCTGTGCGG 4339
Db 2088 GACTTCTTGCTATTTTGGTTTAAAGAGAGCCCATATCAAGAGATG-----CTGCAT 2142
QY 4340 ACGATTTAAATGCTGTTCCGGGCGTACCGTAAACCTGAGAGCTTGTGATCTTTGCCGT 4399
Db 2143 GTGCTATTTAAATTTGTTCCAAATGTCATTAATCTGAGACTGTGATGATTTTTCATTT 2202
QY 4400 TGTCCAGTTCCTCCCAACCAATTTGTGATTTGGGGCTGTCCCTG---CCGTAGAGA 4455
Db 2203 TGTCCAGTTCCTCCCAACCAATTTGTGATTTGGGGCTGTCCCTTACATAGAAATG 2262
QY 4456 CAGAGAGATGGGTGATCTGTTTGAATATGTATGTAGACTGAGCTGATATGAA 4515
Db 2263 CAGAGAGATGGGTGATCTGTTTGAATATGTATGTAGACTGAGCTGATATGAA 2321
QY 4516 GGGGTTATGCTGTCTGTGACATCACTGTAC--TGTGCGCATGTACATCTGTACCG 4574
Db 2322 GTGTTGTGCTGTGTTGTGTGTATCAGCTGTACCTGTGTGAGCATGTATATACCTGT 2381
QY 4575 AAGAGTACT--CTTCTCCATGCGTAAACCAACACCGTGTACAGTCTCATCTACT 4633
Db 2382 CATTAAGAAATTTAGTTCTTTCATGCGAAGCATTAATCTTGTACATGATCTCAATATAT 2441
QY 4634 --GCATTCATTTTACTTTGACAGATGACCTTGTAGACCTGAGGAAAGAC--CCATGTT 4690
Db 2442 TGATTTAATTTATTTTTCACAGATGACCTTGTAGACCATATGAAAGACCTGTGTT 2501
QY 4691 CCGTTGTCTGATGATGATCTAGTGTGCGCGTGTGTTTATTTTCAATCTGCG 4750
Db 2502 TGTTCGCTCAGATTTATCTGTGTGATGATGTTGTTGTTGGGTTTAAATTTTCG 2561
QY 4751 ATGTCTTCCACCATTAATCTAGTAAGACCCCACTGCCCCAGCGTTACATCATATGA 4810
Db 2562 GTGTTTGCATTAACATTAATCTAGTAAGAC--ACCACTGAGGTCGTTAGATCAACGATA 2620
QY 4811 CCACCGTC-----TTAGTCTGCTGTAAGTGAAG--TTTATCCAGTGTCTTTTATGAA 4864
Db 2621 TCCACAGTCTCTTTTATGCTCTGTATCAATGAAGTTTATTCAGTTACTTTTCATGAA 2680
QY 4865 -----TATCTTGAACAGATTAATCTTGTGACAAAGAAAGATGATGAAGTCTCCCTGCA 4919
Db 2681 TGACCTATTTTGAACAGTAAATTTTCTTGACAAAGAAAGATGATGAAGTCTCCCTGCA 2740
QY 4920 ATTAAATTTCCAGTGTTTTACATTTTAACTTGAATCTGTGGGGTGTCTACATTAATAT 4979
Db 2741 ATTAAATTT--CCAATGTTTACATTTTAACTTGAATCTGT--GGAATTTCTTACATTAATAT 2798
QY 4980 GAAATGCGCTCTGCTGCGTGTGATGTTAACTTGTGCTGTAGCTGAAGCCGTGT-- 5037
Db 2799 GAAATGCGCTCTGCTGCGTGTGATGTT--AGATATGCTGTAGCTGAAGCCGTGTGT 2857
QY 5038 -CCTTATATTTAGTTGAAGTCTGGAAGAGAAAT 5071
Db 2858 CTTTAAACATGATGTTGAAGCTCTCAATTAATAT 2892

RESULT 15
AAS84762
ID AAS84762 standard; cDNA; 2874 BP.
XX
AC AAS84762;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #20566.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 4, 2003, 15:24:15 ; Search time 6498.62 Seconds
(without alignments)
19013.884 Million cell updates/sec

Title: US-09-674-237A-1

Perfect score: 5084

Sequence: 1 cggcagcagagagagagtgagag.....gagatcgatatacaagctt 5084

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gse_hum:*
18: em_gse_inv:*
19: em_gse_pln:*
20: em_gse_vrt:*
21: em_gse_fun:*
22: em_gse_mam:*
23: em_gse_mus:*
24: em_gse_pro:*
25: em_gse_rod:*
26: em_gse_pig:*
27: em_gse_vrt:*
28: gb_gse1:*
29: gb_gse2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1598.2	31.4	2126	11	BC020269
2	1582	31.1	2079	11	BC013578
3	809.8	15.9	818	14	CA324450
4	804.2	15.8	906	13	BU519029

Result No.	Score	Query Match	Length	DB ID	Description
5	795.6	15.6	911	13	BQ942708
6	748.8	14.7	2589	11	AK082606
7	727.2	14.3	732	14	CA750495
8	722.8	14.2	725	14	CB248849
9	722.8	14.2	725	12	BM950476
10	721.2	14.2	738	12	BM944544
11	704.8	13.9	729	13	BU704308
12	701.8	13.8	706	14	CB521237
13	699.4	13.8	703	14	CB527154
14	691.2	13.6	713	14	CA329050
15	691.2	13.6	782	12	BG922629
16	684.8	13.5	868	14	CB244555
17	660.2	13.0	896	13	BO941336
18	658.2	12.9	690	13	BO719508
19	656.6	12.9	939	13	BQ179493
20	654.4	12.9	687	14	CB521525
21	646.2	12.7	698	14	CB519152
22	635	12.5	636	14	CD347867
23	633.6	12.5	650	10	BB656585
24	631.4	12.4	1062	13	BO898718
25	625	12.3	636	14	CB526899
26	623	12.3	782	13	BU614534
27	622.6	12.2	694	13	BU054334
28	616	12.1	674	9	AV246405
29	609.6	12.0	616	29	CC325581
30	606.8	11.9	610	13	BO443095
31	602	11.8	603	9	AI594919
32	598	11.8	598	14	CA873077
33	593.2	11.7	606	10	AW910367
34	592.6	11.7	957	14	CB587690
35	587	11.5	600	14	CA528645
36	587	11.5	623	13	BU056026
37	583.8	11.5	587	14	CB724039
38	581.8	11.4	585	14	CB518599
39	581.8	11.4	624	14	BY713771
40	580.4	11.4	582	29	CC183329
41	577.4	11.4	579	14	CA879211
42	572	11.3	572	14	CA895623
43	571.4	11.2	573	14	CA874698
44	566.8	11.1	728	14	BY731854
45	566.4	11.1	734	14	CB269754

ALIGNMENTS

RESULT 1
LOCUS BC020269 2126 bp mRNA
DEFINITION Homo sapiens, clone IMAGE:4899011, mRNA.
ACCESSION BC020269
VERSION BC020269.1 GI:17939664
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 2126)
Strausberg, R.
Direct Submission
Submitted (19-DEC-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.W.A.G.E. Consortium (LINL)
DNA Sequencing by: Genome Sequence Centre,
BC Cancer Agency, Vancouver, BC, Canada
info@bcgsc.bc.ca

Steven Jones, Jennifer Amano, Ian Bosdet, Yaron Butterfield, Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Gai, Leticia Hsiao, Martin Krzyżniński, Reza Kutscher, Oliver Lee, Soo Sen Lee, Victor Ling, Carrie Mathewson, Candice McCreary, Steven Ness, Pawan Pandoh, Anna-Liisa Pribhu, Parvaneh Saedi, Jacquelin Smith, Duane Smalue, Michael Smith, Lorraine Spence, Jeff Stott, Michael Thorne, Miranda Tsai, Natasja van den Bosch, Jill Vardy, George Yang, Scott Zuydamduyn, Marco Marra.

Clone distribution: MCC clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at: <http://image.lnl.gov>
 Series: IRAL Plate: 40 Row: n Column: 1
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 4504796
 This clone has the following problem: no cloning site / microdeletion.

FEATURES	Location/Qualifiers
SOURCE	1. 2126

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/cisone="IMAGE:4899011"
/cisone_type="Pancreas, ep
/cclone_lib="NIH MGC 42"
/lab_host="DH10B-R_"
/notes="Vector: pOT87"

```

Query Match	31.4%;	Score 1598.2;	DB 11;	Length 2126;
Best Local Similarity	85.4%;	Pred. No. 0;		
Matches 1803; Conservative	0;	Mismatches 304;	Indels 4;	Gaps 2;

Oy		92	CCGGGACCGGCCGGAGGACAGGCAGGCGGGCGGGGATGGTGTGCGCGGCTTCGGACT	151
Dd		14	CCGAGCGGGCTCCGGGACGACAAGAGGCGGGCGGGATGGTGTGCCGGCTC	73
Oy		152	CGGCGTTTCCTCGC-GCGGCGTGCGGGCTCACTGATTGTGTGAAGGGCGGCGCGCA	210
Dd		74	CTGCGTCCCTCCAGCGGGCGCGTGAGCGGCACTGATTTGTCCCTGGGGCGGACGGCGA	133
Oy		211	CCCCCGGAGATGAGGCGCTGCATCAGCAAAGTGAACTTAATAGAACCATGGCTCACTTT	270
Dd		134	CCCCCGGAGATGAGGCGCTGCATTAGCAAGTTAAAGTAACAGAACCATGGCTCACTTT	193
Oy		271	CCCAKACCTTTGGSTGTGCTGATGCTGTCTGGCCATACTGTGAGAGAAAGGCCAAG	330
Dd		194	CCAAKACCTTTGGSTGTGAGCCTGGATATCTGGGCCATTACTGTGAGAGAAAGGCCAAG	253
Oy		331	CATGACACAGAGTTCCTTAGCCTGMAAGCCGATAGGGGATTTATTACTGTGTATCAAGCG	390
Dd		254	CATGATCAGAGTTCATAGTTTAAAGCCAAATATCTGAATTCATTACTGTGTATCAAGCT	313
Oy		391	AGGACCTTTTTTCCAATCTGGGTAACTCGAGCCTGACTTAGCACAAATATGGGCGCTA	450
Dd		314	AGAACTTTTTTTCAATCTGGGTAACTCGAGCCTGACTTAGCACAAATATGGGCGCTA	373
Oy		451	GCGGACATGATTAACATGAGAAGATGATCAAGTGATTTTTCCATAGCATTAAGCTT	510
Dd		374	GCTGACATGATTAATATGGAAGATGATCAAGTGAGTTTTCCATAGCATTAAGCTT	433
Oy		511	ATCAAACTGAAGCTACAGAGATATCAGCTCCCCTCACACTTCCCCTGTATGAAAAG	570
Dd		434	ATCAAACTGAAGCTACAGAGATATCAGCTACCCTCGACCTTCCCCTGTATGAAAAG	493
Oy		571	CAACCACTGACTATTTCCAGTGACCAAGCATTTGGTATAGAGGAGATTGTAAGATGCCA	630
Dd		494	CAACCACTGACTATTTCTAGGCGCACAGCATTTGGTATAGGAGGATATGACAGATGCCA	553
Oy		631	CCACTACAGCTGTGCTCTGTGCCAATAGGCTCATTTCAAGTTTGGAAATGTCTTCA	690
Dd		554	CCGCTTACAGCTGTGCTCTCAGTGCCAAAGGATCATTTCAAGTTTGGAAATGTCTTCA	613

QY	691	CCCTTAGATCTTTCTGTCCCTCCAGCAGCAAGTGGCTCCCTGGCTAAAGGGGCTTCCTCC	750
Db	614	ACCTTAGATCTTTCTGTCCCTCCAGCAGCAAGTGGCTCCCTGGCTAAAGGGGCTTCCTCC	673
QY	751	GTCAATACAGCCTCGCCGCTGGTGGGATCCTTCAGACACATAGGCGAAAGTTCTTCC	810
Db	674	GTATATACACTCTTGCTGCACTTGGCTCATCTTCAGCCACATTCGCCAAGAGTTCTTCC	733
QY	811	TTACAGCAATCTGATCCAGGGTCACAATTAAACACTAAGTTACAGAAAGCAATCATTC	870
Db	734	TTTGTAGATCTGGTCCAGGGTCAACAATAACAATAAATTAAGGAGCAGTCAATTT	793
QY	871	GATGTCGCCAGCGCCCTCCAGCAGCAGAAATGGGCTGTGGCTCACTCATCAAGGCTGAA	930
Db	794	GATGGGCGCAATGTGCCCAACAGTGGCAGAGAGGGGCTGTTCTTCAGTCAATCAAGCTGAA	853
QY	921	TACAGGCGATTATTCACAGCGCAGACGACAAACCTATAGTGGACACTTAACAGATCCCGAG	990
Db	854	TACAGCAATTAATTCATATAGTCATGACAAACCTATAGTGGACACTTAACAGATCCCGAG	913
QY	991	GCAAGAACTATATCTCATGCAATCAAGTTTACCCAGGCTCAGCTGGCTTCAATATGAAT	1050
Db	914	GCAAGAACTATATCTTATGCAATCAAGTTTACCAAGGCTCAGCTGGCTTCAATATGAAT	973
QY	1051	CTTTCTGCAATTGATCAAGATGGAAGAACTCACTGCAAGAAATTTATCTAGCTATGAC	1110
Db	974	CTTTCTGCAATTGATCAAGATGGAAGAACTTACACAGAGGAATTTATCTTGGCAATGAC	1033
QY	1111	CTAATATGATGTGGCAATGCTGATCAGCACTGCGCGCCGTCGCGCTCCAGAAATPACATC	1170
Db	1034	CTCATATGATGATGATGTGGCAATCAGCACTGCGCGCACTGCTGCGCTCCAGAAATPACAT	1093
QY	1171	CTCTCTTCTCTCAGAAAGATTCGCTCCGGCAGTGGGATGTCCGTCATTAAGCTCTTCTTCT	1230
Db	1094	CCACTCTCTTTTAAAGAGATTCGATCTGGCAGTGGTATATCTGTCATTAAGCTCAACATCT	1153
QY	1231	GTGGATCAGAGCGCTGCTGAGAGACCGTCTCAAGGATGAGCGACGC--CAGAGAG	1287
Db	1154	GTAGATCAGAGCGCTTACAGAGAGAACAGTTTAAAGATGGAACAACAATTTGAAAG	1213
QY	1288	AAACGCGCTGTGAATTTGAAGATPAAGCGGGGAGAACTTCGACGCGCAGTGTGGAG	1347
Db	1214	AAATTAACCTGTGAAGTTTGAAGATPAAGCGGGGAGAACTTTGAACGTGGCACTTGGAA	1273
QY	1348	CTGGAGAACCGCCGCAAGCGCTCTTGGAGCAGACGCGCAAGAGCAGAGCGGTTGGCT	1407
Db	1274	CTGGAGAAACGAAAGCAAGCAAGCTCTCTGGAAACAGCAGCGCAAGAGCAGAGCGCTGGCC	1333
QY	1408	CAGCTGAGCGCCCGAGCAGAGAGGAAAGCGGGAGCGCCCAAGAGCAGAGCGCCAG	1467
Db	1334	CAGCTGAGCGCGCGAGCAGAGAGGAAAGGAGCGGTGACCCCAAGAGCAAGACGCAAA	1393
QY	1468	CGCAGCTGGAGCTGGAGAAACCACTGTGAAGAGACGCGGAGCTGGAGCGGCAACGAGAG	1527
Db	1394	AGACAACTGGAGACTGGAGAAACAACTGAAAGAGCAGCGGAGCTTGAAGACGCAAGAGAG	1453
QY	1528	GAGGAGAGAGAAAGAGATCGAAGGCGCAGGCGCCGCAAAACGGAACTTGAAGGCGAG	1587
Db	1454	GAGGAGAGAGAAAGAAATTTGAGAGCGAGAGGCTGCMAAACGGGAACCTTGAAGGCGAA	1513
QY	1588	GCACAACTTGAATGGGAAACGGAACCGGAGACAGGAACCTCTGAATCAAGAGAACAGAG	1647
Db	1514	GCACAACTTGAATGGGAAACGGAATTCGAAGCGAAGAACTAATAATCAAGAAACAAAGAA	1573
QY	1648	CAGAGAGGACCGTGTGCTCTGAAGCGAAGAGAAAGTCTGGAGTTTGAATGAAGCT	1707
Db	1574	CAGAGAGCAATATGTTGATCTGAAAGCAAGAAAGAACTTTGGAATTTGAATGAAGCT	1633
QY	1708	CTGAATGCAAAAAGCAATCAGCTGAAGAGAAACCTTCAGATATCAGGTGCGACTGGCA	1767
Db	1634	CTAAATGATTAAGAGCATCACTGAAGGGAACCTTCAAGATATCAGATGTGCAATTGACC	1693
QY	1768	ACCCAGAGGCAAGAAATTTGAGAGCAGAACAGTCTTGAAGAGCTTAAGATTTGCTGAATC	1827

	Query Match	Best local Similarity	Matched 1780;	Conservative	0;	Mismatches	290;	Indels	4;	Gaps	2;
Qy	129	GATGTTGTCGCGGCTCGAGTCTGGGGTTCCTGCG-GCGGCGTGGCGGCTGCATGATT	187								
Db	1	GATGTTGTCGCGGCTCGAGTCTGGGGTTCCTGCGGCTGCATGCGGCGTGAAGCGGCTGATT	60								
Qy	188	TGTGTGAGGGGCGGCGCGCGCACC CGCGAGATGAGCGTTCGATCAGCAAGTGAAC	247								
Db	61	TGTCCCTGGGGCGGCAACGCGGACCCCGCGAGATAGCGGTGATTAGCAAGTGAAC	120								
Qy	248	GTAATAGAACCATGCTCAAGTTCCCAACCTTTCCGATGTAGCGCTGATGCTGGGCCA	307								
Db	121	GTAACGAACCATGCTCAAGTTCCCAACCTTTCCGATGTAGCGCTGATGCTGGGCCA	180								
Qy	308	TAACTGTGAGGAAGGGCCAGATACACAGCTTCCTTAGCTGAAGCCGATAGCGG	367								
Db	181	TAACTGTGAGGAAGGGCCAGATACACAGCTTCCTTAGCTGAAGCCGATATCTG	240								

Db	241	GATTCATTACGGTGATCAAGCTAGAAACCTTTTTTTTCAATCTGGGTACTCACTG	300
Qy	428	TCTTAGCACAATATGCGCCGTAGCGCACTGAATTAACATGAGAGATGATCAAGTG	487
Db	301	TTTATGACAGATATGCGCACTAGCTGCATCATGATATGATGAAAGATGATCAAGTGG	360
Qy	488	AATTTTCATAGGCATAGAGCTTATCAACCTGAAGCTACAAGGATATCAGTCCCTCCA	547
Db	361	AGTTTTCATAGCTATBAAACTTATCAACCTGAAGCTACAAGGATATCAGTCCCTTG	420
Qy	548	CACTTCCCGCTGCATGAAACAGCAACCAAGTGGCTATTTCCAGTCAACAGCATTTGGTA	607
Db	421	CACTTCCCGCTGCATGAAACAGCAACCAAGTGGCTATTTCCAGTCAACAGCATTTGGTA	480
Qy	608	TAGAGGGAATTTGCTAGCATGCCACACTCAACAGCTGTGCTCTGTGCCAATGGGCTCCA	667
Db	481	TGGAGGATATCGCCAGCATGCCACCGCTTACAGCTGTGCTCCAGTCCAAATGGATCCA	540
Qy	668	TTCCAGTTGTGGAAATGTCCACCCCTATATCTTGTCCTCCGACAGCAAGTGGCTC	727
Db	541	TTCCAGTTGTGGAAATGTCTCCAAACCTTAGATCTTGTTCCACAGCAAGTGGCTCC	600
Qy	728	CCCTGGCTAAAGGGGCTCTCCCGTCAATACAGCTCTGCGTGTGGCATCTCTCAG	787
Db	601	CCCTGGCTAAAGGGGCTCCCGCTGTATTAACAACCTCGCTGCAATTTGCTCATCTCGAG	660
Qy	788	CCACATGGCCAAAGAGTTCCTTCTTCCAGCAAGATCTGGTCCAGGGTCACAATTAACACTA	847
Db	661	CCACATGGCCAAAGAGTTCCTTCTTATGATCTGGTCCAGGGTCACAATTAACACTA	720
Qy	848	AGTTACGAAGGGCAATCATTTGATGTGGCAGCGCCCTCCAGACAGCAAGATGGGCTG	907
Db	721	AATTCACAAAGGCACATCTATTGATGTGGCAGGTGCCACCAAGTGGAGTGGGCTG	780
Qy	908	TGCTCAGTCATCAAGGCTGAAATACAGGCAAGTATTCACAAGCCACGACAAAATCTATGA	967
Db	781	TTCTCTCAGTCATCAAGACTGAATATACGGCAATATTCATATGTCAGCAAAAATCTATGA	840
Qy	968	GTGACACTTAAACAGGTCCCGACAGCAAACTATTTCTCATGCAATCAAGTTTACCCAGG	1027
Db	841	GTGACACTTAAACAGGTCCCGACAGCAAACTATTTCTTATGCAATCAAGTTTACCCAGG	900

OY	1028	CTCAGCTGGCTTCAATATGGAATCTTTTGTGACATGATCAAGATGGAACCTCACTGAC	1087
Db	901	CTCAGCTGGCTTCAATATGGAATCTTTTGTGACATGATCAAGATGGAACCTCACTGAC	960
OY	1088	AAGAAATTTATCTTACGTATGCACTTAATGTATGTTGCCATGTCTGGTCAAGCCATGCGCC	1147
Db	961	AGAAATTTATCTTGGCAATGCACTCATTTGATGTATGTCTGGCCAACTGCGCAC	1020
OY	1148	CCGTCCTGGCTCCAAATTAACATCCCTCCCTCCCTCAGAAAGTTGGCTCCGCGAGTGGGA	1207
Db	1021	CTGTCTCCCTCCAAATTAACATCTTCACTTCTTTTGAAGAGTTTGATCTGGCACTGGTA	1080
OY	1208	TGTCGCTCATTAAGCTCTTCTTCTGTGATCAAGAGCTGCTGAGAGCCGTGTCAGAG	1267
Db	1081	TATGTGTATTAAGCTCAACATCTGTAGATCAAGAGCTCAAGAGGAACCAATTTTGAAG	1140
OY	1268	ATGAGCAGCAGC--CAGAGAAAGAACTGCTGTGTACATTTGAGATTAAGACCGGAGA	1322
Db	1141	ATGAACAACAATTAAGAAAAAATTAACCTGTAACTTTGAAATTAAGACCGGAGA	1200
OY	1325	ACTTGACACGAGCGAGTGTGAGCTGTGAGAAAGCGCGCAAGCGCTCTTGAGCAGCAGC	1384
Db	1201	ACTTTGAACGTGGCACTCTGGAACCTGGAGAAACGAAGCGAGCTTCTTGAAACAGCAGC	1266
OY	1385	GCAAAGACAGAGACCGGTTGGCTCAGCTGGAGCGCGCCGAGCAGGAGAGAGAAAGACGGG	1444
Db	1261	GCAAGGACAGAGACCGCTGGCGCCAGCTGGAGCGGCGGAGAGAGAGAGAGAGCGTG	1320
OY	1445	AGCGCCAGAGAGAGAGAGGCAAGCGGCGAGCTGGAGTGGAGAGAGCAGCTGGAGAAACAGC	1500
Db	1321	AGCGCCAGAGAGAGAGAGGCAAGCGGCGAGTGGAGTGGAGAGAGCAGCTGGAGAAAGCAGC	1380
OY	1505	GGGAGCTGAGCGGCGAGCGAGAGAGAGAGAGAGAGAGATCCAGAGGCGCGAGCGCG	1564
Db	1381	GGGAGCTGAGAACCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGCGTG	1440
OY	1565	CAAAACGGGAACTGGAAAGGCAAGCACAATTGAATGGGAACCGAGACAGGAAC	1622
Db	1441	CAAAACGGGAACTTGAAGAGCAACAGCAAACTTGAATGGGAACCGAATCGAAGGCAAGAAC	1500
OY	1625	TCCTGAATCAGAGGAACAAGAGAGCGAGAGGCGACGTGGTCTTGAAGCAAGAGAGAGA	1684
Db	1501	TACTTAATCCAAAGAAACAAAGAACAAAGAGACATAGTGTGTAGAAACAAAGAAAAAG	1566
OY	1685	CTTGGAGTTGAGTTAAGAGCTCTGAATGACAAAGAGCAGTGAAGAGAAACTTC	1744
Db	1561	CTTGGAAATTTGAATTAAGAGCTCTTAATGATTAAGAAACATCAACTTAAGAGGAAACTTC	1620
OY	1745	AGGATATCAGAGTGTGAGCTGGCAACCCAGAGGCAAGAAATGAGAGCGAGCAAGTCTA	1804
Db	1621	AAGATATCAGATGTGAGTTGACACACCAGAGGCAAGAAATGAGAGCGACAAACCAATCTA	1680
OY	1805	GAGAGCTAAGATTTGCTGAAATCACCCACTTACAGACAGCAGTTGACAGAACTTCAGCAA	1864
Db	1681	GAGAGTTGAGATTTGCCGAATACCCACTCTACAGCAACAATTAACAGGAATTCAGCAA	1740
OY	1865	TGCTTGGAAAGACTTATTCAGAGAAACAGATCTCAGTGAACCAATTAAACAAAGTCCAGC	1922
Db	1741	TGCTTGGAAAGACTTATTCAGAGAAACAGATCTCAGTGAACCAATTAACAAAGTTCAGC	1800
OY	1925	AGAAACAGTTTGATAGAGACTGCTGCTTCTTAACCTCAAAAAGAGCTTGGAAAGCAAGAGC	1984
Db	1801	AGAAACAGTTTGACAGAGATTTACCTTGTTCACCTTAAAAAGAGCTTGAAGCAAGAGAC	1866
OY	1985	TGGCCCGGACAGCTCCGGGAGCAGCTGACGAGGTGAGAGAGACCAAGTCAAAAC	2044
Db	1861	TAGCTCGGACAGCACTTACGAGACCAACTGAGTGAAGTGAAGAAAGAACTAGATCAAAAC	1920
OY	2045	TGCAGAGATTTGATGTTTTTCAACAACAGCTGAAGAACTGAGAGAGATTAACAACAAC	2104
Db	1921	TACAGAGATTTGATATTTTCAATATATACGCTGAAGAGATTAAGAGAAATACCAATTAAGC	1980

QY	2105	AGCAACCTCCGAAAGCAAGGTCCTCGAGGCGAGCGACATCGAAGCAGAAAAGCAGGAGA	2164
Db	1981	AACAACCTCCGAAAGCAAAAAGTCATGGAGGCTGAACGATCGAAGCAGAAAAGAACAAAGAC	2040
QY	2165	GGAAAGACCTCGAAGTACAGAAAGCAAAAAGCAAGA	2198
Db	2041	GAAAGATCATGAAATTGAAAAACAAAAAAA	2074

RESULT	3
CALIBRATION	CA324450
LOCUS	
DEFINITION	818 bp mRNA linear EST 26-NOV-2002 U1-M-FYO-CCL-a-05-0-U1.r1 NIH_BMAP_FYO Mus musculus cDNA clone
ACCESSION	CA324450
VERSION	CA324450.1 GI:24542548
KEYWORDS	EST.
SOURCE	Mus musculus (house mouse)
ORGANISM	Mus musculus

REFERENCE	1 (bases 1 to 818)
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/ .
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC).
JOURNAL	Unpublished
COMMENT	Contact: Robert Strausberg, Ph.D.

Tissue procurement: Dr. Jim Lin, University of Iowa
CDNA library preparation: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: MCC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLB at:
<http://image.llnl.gov>
This clone was contributed by the Brain Molecular Anatomy Project (BMAP)
The following repetitive elements were found in this cDNA sequence: 338-455 >(GGA)nSimple_repeat
Seq primer: pIX-5, 5'-GGTCTTTGTTCATCAGGACGCGA-3'

```

FEATURES
SOURCE

Location/Qualifiers
1. 818
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE: 6820974"
/tissue_type="whole brain"
/dev_stage="embryo 13.5,14.5,16.5,17.5dpc"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NH BMAP FY0"
/notes="Organ: Brain; Vector: pYX-Asc; Site_1: Ecor I;
Site_2: Not I; The library was constructed according
to Bonaldi, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured RNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with oligo-dT
primer containing a Not I site. Double strand cDNA was
size selected according to mRNA size fraction, ligated
with Ecor I adaptor, digested with NotI and then cloned
directionally into pYX-Asc vector. The library tag
sequence located between the Not I site and the polyA tail
is AGCGAGACAG. This library was created for the University
Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
Developing Mouse Nervous System', supported by National
Institute of Mental Health (NIMH), Hemlin Chin, Ph.D.,
program coordinator."
238 a 179 c 270 g 127 t 4 others

```

Query Match	15.9%	Score 809.8;	DB 14;	Length 818;
Best Local Similarity	99.3%	Pred. No. 1,16-166;		
Matches 811, Conservative	0;	Mismatches 6;	Indels 0;	Gaps 0;
QY	1087	GAAGAAATTATCCAGCTATGCACCTAATTGATGTTCGACATCTCTCGGTACGCCACTGCGG	1146	

```

Db      2  GAAGATTATCTTACCTATGACCTAATGATGTCATGCTGTCAGCCAGCCG 61
      1147  CCGGTCTGCTGCTCAGAAATCATCCCTCTCTTCAAGAAAGTGTGCTCCGCGACGTGG 1206
      62  CCGGTCTGCTCTCAAAATCATCCCTCTCTTCAAGAAAGTGTGCTCCGCGACGTGG 121
      1207  ATGTCCGTATTAAGCTCTTCTTCTGTGATCAGAGGCTGCTGAGAGCCGTGTGAG 1266
      122  ATGTCCGTATTAAGCTCTTCTTCTGTGATCAGAGGCTGCTGAGAGCCGTGTGAG 181
      1267  GATGACACAGCAGCAGAGAAAGAACTGCTGTGACATTTGAGATTAAGAAACCGAGAAC 1326
      182  GATGACACAGCAGCAGAGAAAGAACTGCTGTGACATTTGAGATTAAGAAACCGAGAAC 241
      1327  TTGAGACGAGGAGGAGTGTGAGCTGAGAAAGCGCGCAGCGCTCTTGAAGACGACCGC 1386
      242  TTGAGACGAGGAGGAGTGTGAGCTGAGAAAGCGCGCAGCGCTCTTGAAGACGACCGC 301
      1387  AAAGACAGAGAGCGGTTGCTCAGCTGAGCGCGCAGAGAGAGAGAAAGCGGAG 1446
      302  AAAGACAGAGAGCGGTTGCTCAGCTGAGCGCGCAGAGAGAGAGAAAGCGGAG 361
      1447  CGCCAGAGCAGAGAGCGCAAGCGGCAAGCTGAGCTGAGAGAGAGCTGAGAGACGCG 1506
      362  CGCCAGAGCAGAGAGCGCAAGCGGCAAGCTGAGCTGAGAGAGAGCTGAGAGACGCG 421
      1507  GAGCTGAGACCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1566
      422  GAGCTGAGACCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 481
      1567  AAAGCGGAACTGGAAGAGCAGCAGCACTTGAATGGGAAACGGAACCGAGACAGAACTC 1626
      482  AAAGCGGAACTGGAAGAGCAGCAGCACTTGAATGGGAAACGGAACCGAGACAGAACTC 541
      1627  CTGAATCAGAGAAACAAGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1686
      542  CTGAATCAGAGAAACAAGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 601
      1687  CTGAGATTGAGATTAAAGCTCTGATGACAAAAGATCAGCTTAAGAGAAACTTCAAG 1746
      602  CTGAGATTGAGATTAAAGCTCTGATGACAAAAGATCAGCTTAAGAGAAACTTCAAG 661
      1747  GATATCAGAGTGTGAGCTGCACTGGCAACCCAGAGAGAGAGAGAGAGAGAGAGAG 1806
      662  GATATCAGAGTGTGAGCTGCACTGGCAACCCAGAGAGAGAGAGAGAGAGAGAGAG 721
      1807  GAGCTAAGAAATGCTGAAATCAACCACTTAAGAGAGAGAGAGAGAGAGAGAGAG 1866
      722  GAGCTAAGAAATGCTGAAATCAACCACTTAAGAGAGAGAGAGAGAGAGAGAGAG 781
      1867  CTTGGAAGACTTATTTCCAGAGAAACGATTAAGAG 1903
      782  CTTGGAAGACTTATTTCCAGAGAAACGATTAAGAG 818

RESULT 4
LOCUS   BUS19029      906 bp      mRNA      linear      EST 12-SEP-2002
DEFINITION  AGENCOURT_10171971 NIH_MGC_134 Mus musculus cDNA clone
IMAGE:6516809 5', mRNA sequence.
ACCESSION  BUS19029
VERSION    BUS19029.1 GI:22826555
KEYWORDS   EST.
SOURCE     Mus musculus (house mouse)
ORGANISM   Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS   NIH-MGC http://mgi.nci.nih.gov/
JOURNAL   National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT   Unpublished
Contact: Robert Strausberg, Ph. D.

```

```

FEATURES
source
  /organism="Mus musculus"
  /mol_type="mRNA"
  /db_xref="taxon:10090"
  /clone="IMAGE:6516809"
  /csize_type="undifferentiated limb"
  /lab_host="DH10B (phage-resistant)"
  /clone_1ib="NIH MGC 134"
  /note="Vector: PCMV-SPORT6.1.cdb; Site 1: EcoRV, Site 2:
  NotI. Cloned unidirectionally. Primer: Oligo dT. Average
  insert size 1.7 Kb. Constructed by Resgen, Invitrogen
  Corp. Note: this is a NIH MGC library."
BASE COUNT      248 a      226 c      235 g      194 t      3 others
ORIGIN
Query Match      15.8%; Score 804.2; DB 13; Length 906;
Best Local Similarity 97.2%; Pred. No. 1.9e-165;
Matches 839; Conservative 0; Mismatches 20; Indels 4; Gaps 2;

      3354  GACGGGAACGATGGGAGCAAGTCGAGATCTTCCCTTAATATATGAGGCTTAAGA 3413
      1  GACGGGAACGATGGGAGCAAGTCGAGATCTTCCCTTAATATATGAGGCTTAAGA 60
      3414  TTGAGAGGCTCTGAACTCTGAGAAACAAGAGAGTTTGAAGAAAAAAGCTGAAATTC 3473
      61  TTGAGAGGCTCTGAACTCTGAGAAACAAGAGAGTTTGAAGAAAAAAGCTGAAATTC 120
      3474  CCAGTTATTGCTTCTTACGCTGCTACTGTGCTCCCAACAATCACTGCTGCTTGGGCA 3533
      121  CCAGTTATTGCTTCTTACGCTGCTACTGTGCTCCCAACAATCACTGCTGCTTGGGCA 180
      3534  GCTGATTCGATCCGGAAGAAAGAACCCAGATGATGATGAGAGAGAGAGAGAGAG 3593
      181  GCTGATTCGATCCGGAAGAAAGAACCCAGATGATGATGAGAGAGAGAGAGAGAG 240
      3594  AAGGAAAAAGCGCCAGATAGAGTGTGTTCCAGCAATTAATGTCAACTTAAGCCCGG 3653
      241  AAGGAAAAAGCGCCAGATAGAGTGTGTTCCAGCAATTAATGTCAACTTAAGCCCGG 300
      3654  AACCAAGCAAAATCAACCCCACTGAGCTACCCAGAGCCGAGAGAGAGAGAGAGAG 3713
      301  AACCAAGCAAAATCAACCCCACTGAGCTACCCAGAGCCGAGAGAGAGAGAGAGAG 360
      3714  GGTGATCGGATGTATGATTAACACCGCCAGAGAGAGAGAGAGAGAGAGAGAGAG 3773
      361  GGTGATCGGATGTATGATTAACACCGCCAGAGAGAGAGAGAGAGAGAGAGAGAG 420
      3774  CCAGATCATCAACGTCCTTCAACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3833
      421  CCAGATCATCAACGTCCTTCAACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
      3834  GCAAGTTGGGCTCTTCCCATTAATTAAGCTGACACAGACATGAGAGAGAGAGAG 3893
      481  GCAAGTTGGGCTCTTCCCATTAATTAAGCTGACACAGACATGAGAGAGAGAGAG 540
      3894  GCAATGATCATATGTTGTCATCCCGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3953
      541  GCAATGATCATATGTTGTCATCCCGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
      3954  TATCCCATATCACTGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4013
      601  TATCCCATATCACTGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660

```

Email: cgaabs-r@mail.nih.gov
 Tissue Procurement: Dr. David Rowe
 cDNA Library Preparation: Invitrogen Corp
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLNL4095 row: 0 column: 18
 High quality sequence stop: 701.

FEATURES

source

Location/Qualifiers
 1. 906
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="IMAGE:6516809"
 /csize_type="undifferentiated limb"
 /lab_host="DH10B (phage-resistant)"
 /clone_1ib="NIH MGC 134"
 /note="Vector: PCMV-SPORT6.1.cdb; Site 1: EcoRV, Site 2:
 NotI. Cloned unidirectionally. Primer: Oligo dT. Average
 insert size 1.7 Kb. Constructed by Resgen, Invitrogen
 Corp. Note: this is a NIH MGC library."
 BASE COUNT 248 a 226 c 235 g 194 t 3 others
 ORIGIN

Query Match 15.8%; Score 804.2; DB 13; Length 906;

Best Local Similarity 97.2%; Pred. No. 1.9e-165;

Matches 839; Conservative 0; Mismatches 20; Indels 4; Gaps 2;

TITLE	High-efficiency full-length cDNA cloning
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)
MEDLINE	99279253
PUBMED	10349636
REFERENCE	
AUTHORS	2
TITLE	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
JOURNAL	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
MEDLINE	Genome Res. 10 (10), 1617-1630 (2000)
PUBMED	11042159
REFERENCE	3
AUTHORS	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishi, K., Kitsuana, T., Tashiro, H., Itoh, M., Suni, N., Ishii, Y., Nakamura, S., Hazama, N., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, S., Sageshima, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunaga, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
TITLE	RIKEN integrated sequence analysis (RISA) system-384-format
JOURNAL	sequencing pipeline with 384 multicapillary sequencer
MEDLINE	Genome Res. 10 (11), 1757-1771 (2000)
PUBMED	20530913
REFERENCE	4
AUTHORS	Kawai, J., Shingawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Aikawa, T., Hara, A., Fukunishi, Y., Kono, H., Adachi, U., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamana, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kankawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, M., Gaasterland, T., Gissi, C., King, B., Kochwa, H., Kusub, P., Lewis, S., Matsuo, Y., Nakaido, I., Pesole, G., Quackenbush, J., Schriml, L.M., Stabili, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M.F., Bromberg, M.J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hotam, M., Hume, D.A., Kamlay, M., Lee, N.H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombere, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seta, T., Shibata, Y., Storch, K.F., Suzuki, H., Toyok, O.K., Wang, K.H., Weitz, C., Whitaker, C., Wilming, L., Wyshaw-Borja, A., Yoshida, K., Haegawa, Y., Kawaji, H., Kohzaki, S. and Hayashizaki, Y.
TITLE	Functional annotation of a full-length mouse cDNA collection
JOURNAL	Nature 409 (6821), 685-690 (2001)
MEDLINE	21085660
PUBMED	11217851
REFERENCE	5
AUTHORS	The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL	Nature 420, 563-573 (2002)
MEDLINE	6 (bases 1 to 2589)
PUBMED	Adachi, J., Aizawa, K., Akiyama, T., Aikawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, M., Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T., Hori, P., Imocani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Kono, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohata, N., Okazaki, Y., Saito, R., Saitoh, C., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shingawa, A., Shiraki, T., Sogabe, Y., Tegen, M., Tagawa, A., Takahashi, F., Takaku-Akashira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, Y., Muramatsu, M. and Hayashizaki, Y.
REFERENCE	Direct Submission
TITLE	Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suenho-cho, Tsunumi-ku, Yokohama,

COMMENT	FEATURES	source
Kanagawa 230-0045, Japan (E-mail:genome-res@gsic.riken.go.jp, URL:htp://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)		
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.		
Please visit our web site for further details.		
URL:htp://genome.gsc.riken.go.jp/		
URL:htp://fantom.gsc.riken.go.jp/.		
Location/Qualifiers		
1..2589		
/organism="Mus musculus"		
/mol_type="mRNA"		
/strain="C57BL/6J"		
/db_xref="FANTOM.DB:C230068L04"		
/db_xref="taxon:10090"		
/clone="C230068L04"		
/tissue_type="cerebellum"		
/clone_id="RIKEN full-length enriched mouse cDNA library"		
/dev_stage="0 day neonate"		
<1..2269		
/note="unnamed protein product; interseccion (SH3 domain protein 1A) (MGDI MGJ:1338069, GB NM_010587, evidence:BLASTN, 99%, match=2267)		
putative"		
/codon_start=2		
/protein_id="DACC8546.1"		
/db_xref="GI:26349813"		
/translation="PRHRGPRKSTSIDTPTESPALKRVASPAARPAITGEERFIM		
YTBSSRQGLTFQGGVIVVTKDGMWMTGDSGVPSNVRILKDSGSITAGS		
TGSGKRRPELQAVIASYAATPEQGLTAPQLILIRKNPGMWGEVLQAKGRQI		
WPNVAVKELASPGTSKITPTELPTKTAQVAPVCQVIGWYDAQNDDELAFSGQI		
INLEKEDPMWKGVSQGVGLFPNSNVKLTITMDPEQCMSPDLHLDMLPTERKQGY		
HELILVTEENEDLOLVATELFOKPLTESELLTEKEVAMIPYMKELIMCNKILKALE		
VRKMSSEKRPVKITGIDILSNQPLPHMPYTRFSCQALGAILIQKTDKADPFEKPRK		
RLANDPRCKMPLSSFTLKPMQVOTRRPLIKNILENTPEHPDHSILKALEABEEL		
CSQVNEGVREKENSDELMEIOAHVCGLEBOLAVNSVTNCLGRKPLHSGKLYKAKS		
NKEIYGLFNDPLLTQITKPLGSSGDKVSPKSNQIKMYKTIPIELNEVILPTFTE		
PSGGEPIFHTSHIDRYVTLAASINERTAWOKIKASLEYITEKKEKREKAVLVRS		
RATGIGRLMNVVGEITLKRCSRHSKSNPICEVYMSGCHITKITIQTLPKRNMSNC		
FFIVRDEQVLCITFERDQLSADDFLGRTIEIRADIKKQSGKVPYKCLLHEVPLP		
GEIVRDLQDLFDEP"		
BASE COUNT 693 a 662 c 688 g 546 t		
ORIGIN		
Query Match 14.7%; Score 748.8; DB 11; Length 2589;		
Beet Local Similarity 99.7%; Pred. No. 3.56-153;		
Matches 750; Conservative 0; Mismatches 2; Indels 0; Gaps 0;		
QY 3148 GGGCCCGTAAGGAATCCACAGCATGATACCTGAGCCCTA CTGAAAGCTCGTAGTCTA 3207		
Db 14 GGGCCCGTAAGGAATCCACAGCATGATACCTGAGCCCTA CTGAAAGCTCGTAGTCTA 73		
QY 3208 AAGAAGTGGCTTCCCGCGCGCCCAAGCCATTCCTCGGAGAAAGATTATTGCATG 3267		
Db 74 AAGAAGTGGCTTCCCGCGCGCCCAAGCCATTCCTCGGAGAAAGATTATTGCATG 133		
QY 3268 TACACATACGAGATTCTGAGCAGAGATTAACTTTTACGACAAAGGAGTGTATTGTG 3327		
Db 134 TACACATACGAGATTCTGAGCAGAGATTAACTTTTACGACAAAGGAGTGTATTGTG 193		
QY 3328 GTTACCAAAGAAAGTGTGACTGGTGTGACGGGAAACGATGGGCGCAAGTCCGAGAGCTCTC 3387		
Db 194 GTTACCAAAGAAAGTGTGACTGGTGTGACGGGAAACGAGAGCTCTC 253		
QY 3388 CCTTCTAACTATGTGAGGCTTTAAAGATTACAGAGGCTCTGGAACTGCTGGGAAAAACAGG 3447		
Db 254 CCTTCTAACTATGTGAGGCTTTAAAGATTACAGAGGCTCTGGAACTGCTGGGAAAAACAGG 313		
QY 3448 AGTTTACGAAAAAACTGAAATTGGCCAGGTTATTGCTTCTTACGCTGCTTACTGTTCC 3507		

Db 314 AGTTGAGAAAAAACTGAAATTGCCAGGTTATTGCTTCTACGCTGCTACTGTCCTCC 373

QY 3508 GAAACAATCACCCTGCTGCTGCGCAGCTGATTTGATTCGGAAAAAGAACCCAGGTGGA 3567

Db 374 GAAACAATCACCCTGCTGCTGCGCAGCTGATTTGATTCGGAAAAAGAACCCAGGTGGA 433

QY 3568 TGGTGGGAAAGGAACTGCAAGCTGAGGAAAAAGCCGAGATGAGGTGGTTTCACGA 3627

Db 434 TGGTGGGAAAGGAACTGCAAGCTGAGGAAAAAGCCGAGATGAGGTGGTTTCACGA 493

QY 3628 AATTATGTCAAACTTTTAAGCCCGGAA CAAGCAAAATCACCCCACTGAGCTACCCAAAG 3687

Db 494 AATTATGTCAAACTTTTAAGCCCGGAA CAAGCAAAATCACCCCACTGAGCTACCCAAAG 553

QY 3688 ACCGAGTGCAGCCGACGAGTGTGCCAGTGTATCGGAGATTAACACCCGCCAGAAC 3747

Db 554 ACCGAGTGCAGCCGACGAGTGTGCCAGTGTATCGGAGATTAACACCCGCCAGAAC 613

QY 3748 GATGACGAACTAGCTTCAGCAAGGCGCAGATCATCAAGTCTCAAGAGAGAGCCG 3807

Db 614 GATGACGAACTAGCTTCAGCAAGGCGCAGATCATCAAGTCTCAAGAGAGAGAGCCG 673

QY 3808 GACTGTGGAAGAGAGAGTCAAGTGGGCAAGTTGGCTCTTCCCATTCATTAATTAAG 3867

Db 674 GACTGTGGAAGAGAGAGTCAAGTGGGCAAGTTGGGCTCTTCCCATTCATTAATTAAG 733

QY 3868 CTGACCAAGACATGAGACCCGACCAAGT 3899

Db 734 CTGACCAAGACATGAGACCCGACCAAGT 765

RESULT 7
CA750495 732 bp mRNA linear EST 27-NOV-2002
LOCUS UI-M-F70-cde-p-10-0-UI.r1 NIH_BMAP_F70 Mus musculus cDNA clone

DEFINITION IMAGE: 6832091 5', mRNA sequence.
CA750495

ACCESSION CA750495.1 GI:25575732
VERSION EST.

KEYWORDS Mus musculus (house mouse)
SOURCE

ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclerognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 732) NIH-MGC http://mgi.mci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgsabers-remail.nih.gov

Tissue Procurement: Dr. Jim Lin, University of Iowa
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa

DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/BLNI at:
http://image.llnl.gov

This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)

The following repetitive elements were found in this cDNA
sequence: 637-722, >(CAG)nSimple_repeat

Seq primer: pyx-5.
Location/Qualifiers

1..732
/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6"

/db_xref="taxon:10090"

/clone="IMAGE: 6832091"

/tissue_type="whole brain"

/dev_stage="embryo 13.5,14.5,16.5,17.5dpc"

/lab_host="DH10B (T1 phage resistant)"

/clone_1ib="NIH BMAP F70"

/note="Organ: Brain; Vector: pyx- Abs; Site_1: Ecor I;

Site 2: Not I: The library was constructed according
Bonaldo, Lemon and Soares, Genome Research, 6:791-806,
1996. Denatured RNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with oligo-dT
primer containing a Not I site. Double strand cDNA was
size selected according to RNA size fraction, ligated
with Ecor I adaptor, digested with NotI and then cloned
directionally into pyx-Abs vector. The library tag
sequence located between the Not I site and the polyA tail
is AGCGAGCAG. This library was created for the University
Iowa Brain Anatomy Project (BMAP): "Gene Discovery in the
Developing Mouse Nervous System", supported by National
Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
program coordinator."

BASE COUNT 197 a 187 c 209 g 139 t
ORIGIN

Query Match 14.3%; Score 727.2; DB 14; Length 732;
Best Local Similarity 99.6%; Pred. No. 1.4e-148;
Matches 729; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 788 CCACATGGCCAAAGAGTTCTTCTTCCAGCAGATGTGTCACAGGTCAATTAACACTA 847

Db 1 CCACATGGCCAAAGAGTTCTTCTTCCAGCAGATGTGTCACAGGTCAATTAACACTA 60

QY 848 AGTTACGAAAGGCGCAATCAATTGATGCGCAGCGCCCTCCAGCAGAGAAATGGGCTG 907

Db 61 AGTTACGAAAGGCGCAATCAATTGATGCGCAGCGCCCTCCAGCAGAGAAATGGGCTG 120

QY 908 TGCCCTGAGTCATCAAGGCTGAAATATACAGGAGTTATTCAACAGCAGACAAAATATGA 967

Db 121 TGCCCTGAGTCATCAAGGCTGAAATATACAGGAGTTATTCAACAGCAGACAAAATATGA 180

QY 968 GTGACACTTAACAGGTCCTCCAGCAGAACTATTCTCAGCATCAAGTTTACCACAG 1027

Db 181 GTGACACTTAACAGGTCCTCCAGCAGAACTATTCTCAGCATCAAGTTTACCACAG 240

QY 1028 CTCAGCTGGCTCAATATGGAATCTTCTGATGATCAAGATGAGAAATCTACTGAG 1087

Db 241 CTCAGCTGGCTCAATATGGAATCTTCTGATGATCAAGATGAGAAATCTACTGAG 300

QY 1088 AAGATTATCTTACTATGACCTTAATGATGTTGTCATGTCGTGACGACTGCGC 1147

Db 301 AAGATTATCTTACTATGACCTTAATGATGTTGTCATGTCGTGACGACTGCGC 360

QY 1148 CCGTCTGCTCCGAATATACCTCTCTTCTTCAAGAGTTGCTCCGCGAGTGGGA 1207

Db 361 CCGTCTGCTCCGAATATACCTCTCTTCTTCAAGAGTTGCTCCGCGAGTGGGA 420

QY 1208 TGTCCGTCATAGCTCTTCTTCTGATGATCAGAGGCTGAGAGGCGTCTGACAGG 1267

Db 421 TGTCCGTCATAGCTCTTCTTCTGATGATCAGAGGCTGAGAGGCGTCTGACAGG 480

QY 1268 ATGAGCAGCAGCAGAGAGAACTGCTGTGACATTTGAAGATGAAGCGGAGAACT 1327

Db 481 ATGAGCAGCAGCAGAGAGAACTGCTGTGACATTTGAAGATGAAGCGGAGAACT 540

QY 1328 TCGAGCGAGGAGTGTGAGCTGAGAGAGCGCCGCAAGCGCTTTGGAGAGCAGCGCA 1387

Db 541 TCGAGCGAGGAGTGTGAGCTGAGAGAGCGCCGCAAGCGCTTTGGAGAGCAGCGCA 600

QY 1388 AAGAGCAGAGCGGTTGGCTCAGCTGAGCGCGCGCAGAGAGAGAAAGCGGAGC 1447

Db 601 AAGAGCAGAGCGGTTGGCTCAGCTGAGCGCGCGCAGAGAGAGAAAGCGGAGC 660

QY 1448 GCCAGAGCAGAGAGCAGAGCGGAGCTGAGAGCTGAGAGAGCAGCTGAGAGAGCAGCGG 1507

Db 661 GCCAGAGCAGAGAGCAGAGCGGAGCTGAGAGCTGAGAGAGCAGCTGAGAGAGCAGCGG 720

QY 1508 AGCTGAGCGGC 1519

Db 721 AGCTGAGCGGC 732

```

RESULT 8
CB248849
LOCUS      725 bp      mRNA      linear      EST 14-FEB-2003
DEFINITION
CB248849
IMAGE: 5719103 5', mRNA sequence.
ACCESSION
CB248849.1
GI:28387663
VERSION
CB248849.1
KEYWORDS
EST.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 725)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgaabs-r@mail.nih.gov
Tissue Procurement: Dr. James Lin, University of Iowa
CDNA Library preparation: Dr. M. Bento Soares, University of Iowa
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)
Seq primer: pyx-5.
FEATURES
source
Location/Qualifiers
1..725
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE: 5719103"
/issue_type="whole brain"
/dev_stage="embryo 15.5 dpc"
/lab_host="DH10B (T1 phage resistant)"
/clone_11b="NIH_BMAP_EXO"
/note="Organ: Brain; Vector: pyx-Asc; Site 1: EcoR I;
Site 2: Not I; The library was constructed according to
Bonaldio, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured mRNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with an
oligo-dT primer containing a Not I site. Double stranded
cDNA was size selected according to mRNA size fraction,
ligated with EcoR I adaptor, digested with Not I, and then
cloned directionally into pyx-Asc vector. The library tag
sequence located between the Not I site and the polyA tail
is GTGGGTGGA. This library was created for the
University of Iowa Mouse Brain Molecular Anatomy Project
(BMAP). 'Gene Discovery in the Developing Mouse Nervous
System', supported by National Institutes of Mental Health
(NIMH), Hemin Chin, Ph.D., program coordinator."
BASE COUNT
250 a 151 c 221 g 102 t 1 others
ORIGIN
Query Match 14.2%; Score 724; DB 14; Length 725;
Best Local Similarity 99.9%; Pred. No. 7.3e-148;
Matches 724; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1594 CTTGATGGGAACGGAACCGGAGACGGAATCTCTGAATCAGAGGAACAGAGCGAGG 1653
DB 1 CTTGATGGGAACGGAACCGGAGACGGAATCTCTGAATCAGAGGAACAGAGCGAGG 60
QY 1654 GGCACCGTGTCTTGAAGCAAGAGAGAGACTCTGAGTTGAGTTAGAGAGCTTGAT 1713
DB 61 GGCACCGTGTCTTGAAGCAAGAGAGAGACTCTGAGTTGAGTTAGAGAGCTTGAT 120
QY 1714 GACAAAGACATCAGTTAGAGAGAAATTCAGAGATATCAGGTGTGACGTGCAACCCAG 1773
DB 121 GACAAAGACATCAGTTAGAGAGAAATTCAGAGATATCAGGTGTGACGTGCAACCCAG 180

```

```

QY 1774 AGCAAGAAATTGAGAGACGGAACAGTCTAGAGAGCTAAGAATTGCTGAATCACCAC 1833
DB 181 AGCAAGAAATTGAGAGACGGAACAGTCTAGAGAGCTAAGAATTGCTGAATCACCAC 240
QY 1834 TTACAGCAGCAGTTCAGAGAAATCTCAGCAAAATGCTTGAAGAATTATTCAGAGAAACAG 1893
DB 241 TTACAGCAGCAGTTCAGAGAAATCTCAGCAAAATGCTTGAAGAATTATTCAGAGAAACAG 300
QY 1894 ATACTCAGTGCACAGTTAAACAAAGTCCAGCAAAACAGTTGCATAGAGACTGCTTCTT 1953
DB 301 ATACTCAGTGCACAGTTAAACAAAGTCCAGCAAAACAGTTGCATAGAGACTGCTTCTT 360
QY 1954 ACCCTCAAGAGCCTTGGAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2013
DB 361 ACCCTCAAGAGCCTTGGAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
QY 2014 GACGAGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2073
DB 421 GACGAGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
QY 2074 CTGAAGGAACTGAGAGAGATACATACGCAACAGCACTTCAGAGAGAGAGAGAGAGAG 2133
DB 481 CTGAAGGAACTGAGAGAGATACATACGCAACAGCACTTCAGAGAGAGAGAGAGAGAG 540
QY 2134 GCGAGGCGCAGTGAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2193
DB 541 GCGAGGCGCAGTGAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
QY 2194 GAAGAGCGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2253
DB 601 GAAGAGCGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
QY 2254 GAGAGACAGCCACGCGCCCGGAAACCCACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2313
DB 661 GAGAGACAGCCACGCGCCCGGAAACCCACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
QY 2314 GTCCAG 2318
DB 721 GTCCAG 725

```

```

RESULT 9
BM950476
LOCUS      732 bp      mRNA      linear      EST 14-MAR-2002
DEFINITION
BM950476
IMAGE:5687260 5', mRNA sequence.
ACCESSION
BM950476
VERSION
BM950476.1
GI:19434066
KEYWORDS
EST.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 732)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgaabs-r@mail.nih.gov
Tissue Procurement: Dr. James Lin, University of Iowa
CDNA Library preparation: Dr. M. Bento Soares, University of Iowa
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)
Seq primer: pyx-5.
FEATURES
source
Location/Qualifiers
1..732
/organism="Mus musculus"

```

```

/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:5687260"
/tissue_type="whole brain"
/dev_stage="embryo 18.5 dpc"
/lab_host="DH10B (T1 phage resistant)"
/clone_id="NIH BMAP EHO"
/note="Organ: brain; Vector: pYX-Asc; Site 1: Ecor I;
Site 2: Not I; The library was constructed according to
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured mRNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with an
oligo-dT primer containing a Not I site. Double stranded
cDNA was size selected according to mRNA size fraction,
ligated with Ecor I adaptor, digested with Not I, and then
cloned directionally into pYX-Asc vector. The library tag
sequence located between the Not I site and the polyA tail
, is CAGCCACGAC. This library was created for the
University of Iowa Mouse Brain Molecular Anatomy Project
(BMAP): 'Gene Discovery in the Developing Mouse Nervous
System', supported by National Institute of Mental Health
(NIMH), Hemin Chin, Ph.D., program coordinator."
BASE COUNT      208 a      172 c      201 g      148 t      3 others
ORIGIN
Query Match      14.2% Score 722.8; DB 12; Length 732;
Best Local Similarity 99.3%; Pred. No. 1.3e-147;
Matches 724; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 3075 AGACATGTGTGGTGGAGAAATTCAGAGGAGGTTGTTCCCAAGTCTTACGT 3134
DB 4 ACATNNGTGTTGGTGGAGAAATTCAGAGGAGGTTGTTCCCAAGTCTTACGT 63
QY 3135 GAAACTCATTTAGAGGCGCCGTGAGAAATTCAGAGGAGGTTGTTCCCAAGTCTTACGT 3194
DB 64 GAAACTCATTTAGAGGCGCCGTGAGAAATTCAGAGGAGGTTGTTCCCAAGTCTTACGT 123
QY 3195 TCCTGCTGTGTTAAGAGAGTGGCTTCCCGGCGCCCAAGCCAGCATTTCCCGAGAAAG 3254
DB 124 TCCTGCTGTGTTAAGAGAGTGGCTTCCCGGCGCCCAAGCCAGCATTTCCCGAGAAAG 183
QY 3255 GTTTATTCATGTACATACAGAGGTTCTGAGCAAGAGATTTAACTTTACAGCAAG 3314
DB 184 GTTTATTCATGTACATACAGAGGTTCTGAGCAAGAGATTTAACTTTACAGCAAG 243
QY 3315 GGATGTGATGTGTGTTACCAAGAAAGATGTGACTGTGAGCGGAAAGCGTGGCGCA 3374
DB 244 GGATGTGATGTGTGTTACCAAGAAAGATGTGACTGTGAGCGGAAAGCGTGGCGCA 303
QY 3375 GTCCGAGTCTTCCCTTAACTATGTAGAGGTTAAAGTTCAAGGGCTCTGGAATGC 3434
DB 304 GTCCGAGTCTTCCCTTAACTATGTAGAGGTTAAAGTTCAAGGGCTCTGGAATGC 363
QY 3435 TGGGAAAACAGGAGATTTAGAAAAAACTGAATTTGCCAGTTATTTGCTTCAAGC 3494
DB 364 TGGGAAAACAGGAGATTTAGAAAAAACTGAATTTGCCAGTTATTTGCTTCAAGC 423
QY 3495 TGTACTGTGTCCGAAACAATCACCCTGCTCTGAGGAGCTGATTTGATCCGAAAA 3554
DB 424 TGTACTGTGTCCGAAACAATCACCCTGCTCTGAGGAGCTGATTTGATCCGAAAA 483
QY 3555 GAAACCCAGTGTGTGAGGAGGAACTGCAAGCTGAGGAGAAAGCCGCAATGAG 3614
DB 484 GAAACCCAGTGTGTGAGGAGGAACTGCAAGCTGAGGAGAAAGCCGCAATGAG 543
QY 3615 GTGTTTCCAGCAATTAATGCAAACTTAAAGCCCGGAAACAACAATCAGCCCAAC 3674
DB 544 GTGTTTCCAGCAATTAATGCAAACTTAAAGCCCGGAAACAACAATCAGCCCAAC 603
QY 3675 TGAGCTACCCAGAGCGAGTGCAGCAGCAAGTGTGCAGGTGATCGGAGTGTACGATTA 3734
DB 604 TGAGCTACCCAGAGCGAGTGCAGCAGCAAGTGTGCAGGTGATCGGAGTGTACGATTA 663

```

```

QY 3735 CAGCCGCCAGAGATGACGAACTAGCCTTACGCAAGGCCAGATCATCAAGTCTTCAA 3794
DB 664 CAGCCGCCAGAGATGACGAACTAGCCTTACGCAAGGCCAGATCATCAAGTCTTCAA 723
QY 3795 CAGAGAGGA 3803
DB 724 CAGAGAGGA 732
RESULT 10
LOCUS BM944544
DEFINITION UI-M-EHO-bv-r-j-08-0-UI.r1 NIH BMAP_EHO Mus musculus cDNA clone
ACCESSION BM944544
VERSION BM944544.1 GI:19428129
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 738)
REFERENCE NIH-MGC http://mgi.nci.nih.gov/
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished
COMMENT Contact: Robert Strauberg, Ph.D.
Email: gahbs-romail.nih.gov
Tissue Procurement: Dr. James Lin, University of Iowa
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)
Seq primer: pYX-5.
FEATURES
location/Qualifiers
1..738
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:5687260"
/tissue_type="whole brain"
/dev_stage="embryo 18.5 dpc"
/lab_host="DH10B (T1 phage resistant)"
/clone_id="NIH BMAP EHO"
/note="Organ: brain; Vector: pYX-Asc; Site 1: Ecor I;
Site 2: Not I; The library was constructed according to
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured mRNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with an
oligo-dT primer containing a Not I site. Double stranded
cDNA was size selected according to mRNA size fraction,
ligated with Ecor I adaptor, digested with Not I, and then
cloned directionally into pYX-Asc vector. The library tag
sequence located between the Not I site and the polyA tail
, is CAGCCACGAC. This library was created for the
University of Iowa Mouse Brain Molecular Anatomy Project
(BMAP): 'Gene Discovery in the Developing Mouse Nervous
System', supported by National Institute of Mental Health
(NIMH), Hemin Chin, Ph.D., program coordinator."
BASE COUNT      198 a      213 c      184 g      143 t
ORIGIN
Query Match      14.2% Score 721.2; DB 12; Length 738;
Best Local Similarity 99.5%; Pred. No. 3e-147;
Matches 734; Conservative 0; Mismatches 3; Indels 1; Gaps 1;
QY 2568 CCAGACTGAGAGCCAGAGTGTGAGAGAGACCTGAAGAGAACGGAGTGTCC 2627

```

Db 478 GTTCCATGTCGTGTCACGCACTGCGCCGCTCCTGCTCCAGATAATACCTCTCTCC 537
 QY 1180 TTGAGAAAGTTGCTCGCGAGTGGATGTCCTGATTAACCTCTTCTGTGTGATCAG 1239
 Db 538 TTCGAAAGATTCGTCGCGAGTGGATGTCCTGATTAACCTCTTCTGTGTGATCAG 597
 QY 1240 AGGCTGCTGAGGACGCTGCTGAGAGATGAGCAGCAGAGAAAGAACTGCTGTG 1299
 Db 598 AGGCTGCTGAGGACGCTGCTGAGAGATGAGCAGCAGAGAAAGAACTGCTGTG 657
 QY 1300 ACATTGTAAGATTAAGACCGGAGAACTTCGAGCAGGACAGTGTGAGCTGAGAGCGC 1359
 Db 658 ACATTGTAAGATTAAGACCGGAGAACTTCGAGCAGGACAGTGTGAGCTGAGAGAGCGC 717
 QY 1360 CGCCA.1364
 Db 718 CGCAA 722

RESULT 12
 CBS21237 706 bp mRNA linear EST 28-MAR-2003
 LOCUS UI-M-GH0-cem-g-10-0-UI.r1 NIH_BMAP_GH0 Mus musculus cDNA clone
 DEFINITION IMAGE: 6841475 5', mRNA sequence.
 ACCESSION CBS21237
 VERSION CBS21237.1 GI:29354592
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 706)
 NIH-MGC http://mgc.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: rgs@bhs-rcmail.nih.gov
 Tissue Procurement: Dr. Jim Lin, University of Iowa
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Distribution information can be found at
 http://genome.uiowa.edu/distribution/mousef1.html
 This clone was contributed by the Brain Molecular Anatomy Project
 (BMAP)
 The following repetitive elements were found in this cDNA
 sequence: 220-337, >(GGA)n#Simple_repeat
 Seq primer: PYX-5.

FEATURES
 source location/Qualifiers
 1. 706
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /clone="IMAGE: 6841475"
 /tissue_type="Whole brain"
 /dev_stage="1, 5, and 15 days newborn"
 /lab_host="DH10B (TI phage resistant)"
 /clone_lib="NIH BMAP GH0"
 /note="Organ: Brain; Vector: pYX-Asc; Site: 1: Ecor I;
 Site 2: Not I; The library was constructed according
 Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
 1996. Denatured RNA was size fractionated on a 1% agarose
 gel. First strand cDNA synthesis was primed with oligo-dT
 primer containing a Not I site. Double strand cDNA was
 size selected according to mRNA size fraction, ligated
 with Ecor I adaptor, digested with NotI and then cloned
 directionally into pYX-Asc vector. The library tag
 sequence located between the Not I site and the polyA tail
 is CGAAGTGAAT. This library was created for the University
 Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
 Developing Mouse Nervous System', supported by National
 Institute of Mental Health (NIMH), Hemn Chin, Ph.D.,

BASE COUNT 219 a 143 c 247 g 96 t 1 others
 ORIGIN
 Query Match 13.8%; Score 701.8; DB 14; Length 706;
 Best Local Similarity 99.6%; Pred. No. 5,4e-143;
 Matches 703; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1204 GGGATGTCCTGATTAAGCTTTCTTCTGTGTGATTCAGAGGCTGCTGAGAGCCGCTCTCA 1263
 Db 1 GGGATGTCCTGATTAAGCTTTCTTCTGTGTGATTCAGAGGCTGCTGAGAGCCGCTCTCA 60
 QY 1264 GAGGATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1323
 Db 61 GAGGATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 120
 QY 1324 AACTTGAGCGAGGACAGTGTGAGCTGAGAGAGCGCCGCAAGCGCTCTTGAGCAGCAG 1383
 Db 121 AACTTGAGCGAGGACAGTGTGAGCTGAGAGAGCGCCGCAAGCGCTCTTGAGCAGCAG 180
 QY 1384 CGCAAAAGCAGAGCGCGCTTGCTCAGCTGAGAGCGCGCCGCAAGCGCTCTTGAGCAGCAG 1443
 Db 181 CGCAAAAGCAGAGCGCGCTTGCTCAGCTGAGAGCGCGCCGCAAGCGCTCTTGAGCAGCAG 240
 QY 1444 GAGCGCCAGAGCAGAGCGCGCCGCAAGCGCTTGCTCAGCTGAGAGCGCGCCGCAAGCGCT 1503
 Db 241 GAGCGCCAGAGCAGAGCGCGCCGCAAGCGCTTGCTCAGCTGAGAGCGCGCCGCAAGCGCT 300
 QY 1504 CGGAGCTGAGCGCGCCGCAAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1563
 Db 301 CGGAGCTGAGCGCGCCGCAAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
 QY 1564 GCAGAAAGCGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1623
 Db 361 GCAGAAAGCGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
 QY 1624 CTCCTGATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1683
 Db 421 CTCCTGATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
 QY 1684 ACTCTGAGAGTTGAGTTAGAGCTCTGATGACAAAAGATCAGTGAAGAGAACTT 1743
 Db 481 ACTCTGAGAGTTGAGTTAGAGCTCTGATGACAAAAGATCAGTGAAGAGAACTT 540
 QY 1744 CAGGATATCAGGTGTGAGCTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGT 1803
 Db 541 CAGGATATCAGGTGTGAGCTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGT 600
 QY 1804 AGAGAGCTAAGAAATTTGCTGAATCACCACCTTACAGCAGCAGTGAAGTTCAGCAA 1863
 Db 601 AGAGAGCTAAGAAATTTGCTGAATCACCACCTTACAGCAGCAGTGAAGTTCAGCAN 660
 QY 1864 ATGCTTGAAGACTTATTCAGAGAAACAGATCTCAGTGAACAGT 1909
 Db 661 ATGCTTGAAGACTTATTCAGAGAAACAGATCTCAGTGAACAGT 706

RESULT 13
 CBS27154 703 bp mRNA linear EST 28-MAR-2003
 LOCUS UI-M-FY0-cfh-m-16-0-UI.r1 NIH_BMAP_FY0 Mus musculus cDNA clone
 DEFINITION IMAGE: 6849689 5', mRNA sequence.
 ACCESSION CBS27154
 VERSION CBS27154.1 GI:29360627
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 703)
 NIH-MGC http://mgc.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished


```

Db      1  CCAAGTGAAGAGCAGAGATGCTTGGAGAGAGCTGAAAAGGAGACGGAGATGTTCCC 60
Qy      2628 TGCAGAACTATGCAAGAAAGATTCAGAGAAATGAGGTTCCACTCCAGCCAAACAGTAGAC 2687
Db      61  TGCAGAACTATGCAAGAAAGATTCAGAGAAATGAGGTTCCACTCCAGCCAAACAGTAGAC 120
Qy      2688 CGATCTGACATCTGCCCCCTGCCCCCAAACTGGCTCTGCGTAGAACCCCTGCTCTTGGC 2747
Db      121  CGATCTGACATCTGCCCCCTGCCCCCAAACTGGCTCTGCGTAGAACCCCTGCTCTTGGC 180
Qy      2748 AGTGAACCTCTTCTGAGGCCCCCTCCACAAACCCCAACAACTGGGCGAGCTTCACTTCACAGT 2807
Db      181  AGTGAACCTCTTCTGAGGCCCCCTCCACAAACCCCAACAACTGGGCGAGCTTCACTTCACAGT 240
Qy      2808 GCCCAGAGAGCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2867
Db      241  GCCCAGAGAGCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
Qy      2868 TTCTCTGACCGTACCTAGTCTGCGCACTTACCGGAGATCAGCCCTTACCCAGCCAC 2927
Db      301  TTCTCTGACCGTACCTAGTCTGCGCACTTACCGGAGATCAGCCCTTACCCAGCCAC 360
Qy      2928 AGCAGCTGCTCTCTCCCATCTCCCTCTGCTGCGCAGGGTGAAGAGAGAGAGAGAGAGAGAGAGAG 2987
Db      361  AGCAGCTGCTCTCTCCCATCTCCCTCTGCTGCGCAGGGTGAAGAGAGAGAGAGAGAGAGAGAGAGAG 420
Qy      2988 AGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3047
Db      421  AGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
Qy      3048 TGAAGTATCAGCCGTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3107
Db      481  TGAAGTATCAGCCGTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540
Qy      3108 GAAGGTTGCTTCCCAAGTCTTACGTAAGTCTTACGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3167
Db      541  GAAGGTTGCTTCCCAAGTCTTACGTAAGTCTTACGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
Qy      3168 AACGATGATCTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3227
Db      601  AACGATGATCTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
Qy      3228 CGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3287
Db      661  CGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 719
Qy      3288 GCAGAGAGATTTAAGCTT 3305
Db      720  GCAGAGAGATTTAAGCTT 737

RESULT 11
LOCUS   BU704308 729 bp mRNA linear EST 09-OCT-2002
DEFINITION
UI-M-FOO-bzs-j-21-0-UI.r1 NIH BMAP_FOO Mus musculus cDNA clone
IMAGE: 6406772 5', mRNA sequence.
ACCESSION
BU704308
VERSION
BU704308.1 GI:23632308
KEYWORDS
EST.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 729)
AUTHORS
NIH-MGC http://mgi.nci.nih.gov/.
TITLE
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: rgs@nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

```

```

FEATURES
source
    Clone Distribution: MGC clone distribution information can be
    found through the I.M.A.G.E. Consortium/LNL at:
    http://image.lnl.gov
    This clone was contributed by the Brain Molecular Anatomy Project
    (BMAP)
    Seq primer: pYX-5.
    Location/Qualifiers
        1..729
            /organism="Mus musculus"
            /mol_type="mRNA"
            /strain="C57BL/6"
            /db_xref="taxon:10090"
            /clone="IMAGE: 6406772"
            /tissue_type="whole brain"
            /dev_stage="embryo 12.5dpc"
            /lab_host="DH10B (T1 phase resistant)"
            /clone_11b="NIH BMAP_F00"
            /note="Organ: Brain; Vector: pYX-Arc; Site 1: EcoR I;
            Site 2: Not I; The library was constructed according
            Bonaldi, Lennon and Soares, Genome Research, 6:791-806,
            1996. Denatured RNA was size fractionated on a 1% agarose
            gel. First strand cDNA synthesis was primed with oligo-dT
            primer containing a Not I site. Double strand cDNA was
            size selected according to mRNA size fraction, ligated
            with EcoR I adaptor, digested with NotI and then cloned
            directionally into pYX-Arc vector. The library tag
            sequence located between the Not I site and the polyA tail
            is TAGAGAGGCC. This library was created for the University
            Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
            Developing Mouse Nervous System', supported by National
            Institute of Mental Health (NIMH), Hemlin Chin, Ph.D.,
            program coordinator."
BASE COUNT
176 a 206 c 174 g 172 t 1 others
ORIGIN
Query Match 13.9%; Score 704.8; DB 13; Length 729;
Best Local Similarity 99.2%; Pred. No. 1.2e-143;
Matches 719; Conservative 0; Mismatches 3; Indels 3; Gaps 1;
Qy      640 GCTGTTGCTCTGTCGCAATGAGGCTTCATTCAGATTGTTGAATGCTTCACCTTTAGTA 699
Db      1  GCTGTTGCTCTGTCGCAATGAGGCTTCATTCAGATTGTTGAATGCTTCACCTTTAGTA 60
Qy      700 TCTTCGTCCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 759
Db      61  TCTTCGTCCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
Qy      760 CCTGTCCTGAGCTTGGCATCTGCAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 819
Db      121  CCTGTCCTGAGCTTGGCATCTGCAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 177
Qy      820 TCTGTCAGAGGTCACATTAACCTAAGTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 879
Db      178  TCTGTCAGAGGTCACATTAACCTAAGTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 237
Qy      880 AGGCGCCCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 939
Db      228  AGGCGCCCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 297
Qy      940 TTATTCAACAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 999
Db      228  TTATTCAACAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 357
Qy      1000 ATTCTCATGCAATCAAGTTTAAACCCAGGCTGAGCTGCTCAATATGGAATCTTTCTGAC 1059
Db      358  ATTCTCATGCAATCAAGTTTAAACCCAGGCTGAGCTGCTCAATATGGAATCTTTCTGAC 417
Qy      1060 ATTGATCAAGATGAGAAATCACTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1119
Db      418  ATTGATCAAGATGAGAAATCACTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 477
Qy      1120 GTTGCATGTCGTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1179

```

COMMENT

Contact: Robert Strausberg, Ph.D.
 Email: cga@bbs-remail.nih.gov
 Tissue Procurement: Dr. Jim Lin, University of Iowa
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Distribution information can be found at
<http://genome.uiowa.edu/distribution/mousefl.html>
 This clone was contributed by the Brain Molecular Anatomy Project (BMAP)

FEATURES

source

Seq primer: PYX-5.

Location/Qualifiers

1..703

/organism="Mus musculus"

/strain="CS7BL/6"

/db_xref="taxon:10090"

/clone="IMAGE: 6849689"

/tissue_type="whole brain"

/dev_stage="embryo 13.5,14.5,16.5,17,5dpc"

/lab_host="DH10B (T1 phage resistant)"

/clone_lib="NIH_BMAP_FY0"

/note="Organ: Brain; Vector: PYX-Asc; Site: 1: Ecor I; Site: 2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with Ecor I adaptor, digested with NotI and then cloned directionally into PYX-Asc vector. The library tag is AGCGAGCAG. This library was created for the University of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemin Chin, Ph.D., program coordinator."

BASE COUNT 228 a 154 c 225 g 94 t 2 others
 ORIGIN

Query Match 13.8%; Score 699.4; DB 14; Length 703;
 Best Local Similarity 99.6%; Pred. No. 1.8e-142;
 Matches 700; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

1957 CTCAAAAGAGCTTGGAGCAAGAGAGCTGCGCCGACAGCTCCGGAGAGAGCTGGAC 2016
 1 CTCAAAAGAGCTTGGAGCAAGAGAGCTGCGCCGACAGCTCCGGAGAGAGCTGGAC 60
 2017 GAGGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2076
 61 GAGGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
 2077 AAGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2136
 121 AAGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
 2137 GCGGCACTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2196
 181 GCGGCACTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240
 2197 GAGGCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2256
 241 GAGGCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
 2257 GAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2316
 301 GAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
 2317 AAGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2376
 361 AAGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
 2377 TTCGATCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2436

Db 421 TTCGATCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
 2437 AAAGCCCGCTTACCATTTCTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2496
 Db 481 AAAGCCCGCTTACCATTTCTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540
 2497 TACCCCTTGAATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2556
 541 TACCCCTTGAATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
 2557 GTGATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2616
 601 GTGATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
 2617 GGATGTTCCCTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2659
 661 GGATGTTCCCTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 703

RESULT 14

CA329050

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

CA329050 713 bp mRNA linear EST 27-NOV-2002
 UI-M-FY0-cdc-j-23-0-UI.r1 NIH_BMAP_FY0 Mus musculus cDNA clone
 IMAGE: 6827736 5', mRNA sequence.
 CA329050.1 GI:24547148
 EST.
 Mus musculus (house mouse)
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sclerogonathi; Muridae; Murinae; Mus.
 1 (bases 1 to 713)
 NIH-MGC <http://mgi.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished
 Contact: Robert Strausberg, Ph.D.
 Email: cga@bbs-remail.nih.gov
 Tissue Procurement: Dr. Jim Lin, University of Iowa
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/BLM at:
<http://image.llnl.gov>
 This clone was contributed by the Brain Molecular Anatomy Project (BMAP)

The following repetitive elements were found in this cDNA
 sequence: 341-458, >(GGA)n\$imple_repeat
 Seq primer: PYX-5.

FEATURES

source

1..713

/organism="Mus musculus"

/mol_type="mRNA"

/strain="CS7BL/6"

/db_xref="taxon:10090"

/clone="IMAGE: 6827736"

/tissue_type="whole brain"

/dev_stage="embryo 13.5,14.5,16.5,17,5dpc"

/lab_host="DH10B (T1 phage resistant)"

/clone_lib="NIH_BMAP_FY0"

/note="Organ: Brain; Vector: PYX-Asc; Site: 1: Ecor I; Site: 2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with Ecor I adaptor, digested with NotI and then cloned directionally into PYX-Asc vector. The library tag is AGCGAGCAG. This library was created for the University of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the


```

Db      541  TCACTGCGTCAGAGCGAGAAATGCAAGAGGTGGTCTTTGG 600
Qy      4166 GGGTTTCCCTAGTCACTGAGCTGACCGGCCCCGCTTCAACGGGCGCTTCAATAGTT 4225
Db      601  GGGTTTCCCTAGTCACTGAGCTGACGG--CCGCTTCACACGGGCGCTTCAATAGTT 657
Qy      4226 TAAGATTATTTTAAATGATATTTAGCCTTTTAAATAAATCTCAATCAATTACTTCT 4285
Db      658  TAAGATTATTTTAAATGATATTTAGCCTTTTAAATCAATCAATTACTTCT 716
Qy      4286 TTGCTATTTTGGTTTAC--AAAAACACCCACTATCAAGAGTGC 4329
Db      717  TTGCGTATATGTTTTTACCAAAAAGACCACTATCAAGAGTGC 762

```

Search completed: December 5, 2003, 00:39:52
 Job time : 6505.62 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 4, 2003, 15:25:00 ; Search time 220.815 Seconds
(without alignments)
10167.293 Million cell updates/sec

Title: US-09-674-237A-1

Perfect score: 5084
Sequence: 1 cggcagcagagagagtgagag.....gagaatcgatcatcaagctt 5084

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

Issued Patents NA: *
1: /cgn2_6/prodata/2/ina/5A COMB.seq: *
2: /cgn2_6/prodata/2/ina/5B COMB.seq: *
3: /cgn2_6/prodata/2/ina/6A COMB.seq: *
4: /cgn2_6/prodata/2/ina/6B COMB.seq: *
5: /cgn2_6/prodata/2/ina/PTUS COMB.seq: *
6: /cgn2_6/prodata/2/ina/backfile1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	484.8	9.5	2017	US-09-404-879A-72	Sequence 72, App1
2	484.8	9.5	2017	US-09-338-933-72	Sequence 72, App1
3	484.8	9.5	2017	US-09-215-681-72	Sequence 72, App1
4	333.2	6.6	2873	US-08-630-915A-193	Sequence 193, App
5	258	5.1	747	US-08-630-915A-39	Sequence 39, App1
6	151.4	3.0	531	US-09-404-879A-5	Sequence 5, App1
7	151.4	3.0	531	US-09-338-933-5	Sequence 5, App1
8	151.4	3.0	531	US-09-215-681-5	Sequence 5, App1
9	144	2.8	480	US-09-404-879A-60	Sequence 60, App1
10	144	2.8	480	US-09-338-933-60	Sequence 60, App1
11	144	2.8	480	US-09-215-681-60	Sequence 60, App1
12	144	2.8	3489	US-08-728-323A-1	Sequence 1, App1
13	144	2.8	3489	US-09-298-568-1	Sequence 1, App1
14	144	2.8	3489	US-09-410-399-1	Sequence 1, App1
15	144	2.8	32207	US-08-770-379-20	Sequence 20, App1
16	144	2.8	32207	US-08-757-659A-20	Sequence 20, App1
17	144	2.8	32207	US-09-230-371A-20	Sequence 20, App1
18	121.6	2.4	531	US-09-404-879A-4	Sequence 4, App1
19	121.6	2.4	531	US-09-338-933-4	Sequence 4, App1
20	119.6	2.4	9551	US-08-056-200-93	Sequence 93, App1
21	119.6	2.4	9551	US-08-800-644-93	Sequence 93, App1
22	119.2	2.3	1926	US-09-249-585A-2	Sequence 2, App1
23	119.2	2.3	1926	US-09-410-399-3	Sequence 3, App1
24	119.2	2.3	2580	US-09-050-863-2	Sequence 2, App1
25	119.2	2.3	2580	US-09-359-081-2	Sequence 2, App1
26	119.2	2.3	5452	US-09-130-114-1	Sequence 1, App1
27	119.2	2.3	5452	US-09-130-114-1	Sequence 1, App1

C	28	119.2	2.3	8705	4	US-09-647-344A-14	Sequence 14, App1
	29	119.2	2.3	9600	3	US-08-910-647-1	Sequence 1, App1
	30	119.2	2.3	9600	4	US-09-620-925-1	Sequence 1, App1
	31	119.2	2.3	10596	1	US-07-884-811-15	Sequence 15, App1
	32	119.2	2.3	10596	1	US-07-885-971-15	Sequence 15, App1
	33	119.2	2.3	10596	1	US-08-087-783A-15	Sequence 15, App1
	34	119.2	2.3	10596	1	US-08-194-088B-15	Sequence 15, App1
	35	119.2	2.3	10596	2	US-08-194-088B-15	Sequence 15, App1
	36	119.2	2.3	10596	5	PCT-US93-04648-15	Sequence 15, App1
	37	107	2.1	7218	1	US-08-232-463-14	Sequence 14, App1
	38	95.8	1.9	16442	1	US-08-781-891-208	Sequence 208, App
	39	95.8	1.9	16442	4	US-09-618-166-208	Sequence 208, App
	40	87.4	1.7	397	3	US-09-253-691-3	Sequence 3, App1
	41	87.4	1.7	51259	3	US-08-781-891-209	Sequence 209, App
	42	87.4	1.7	51259	4	US-09-618-166-209	Sequence 209, App
	43	86.2	1.7	543	6	5273901-6	Sequence 3, App1
	44	83.6	1.6	1037	4	US-09-181-585-3	Sequence 3, App1
	45	83.6	1.6	1159	4	US-09-181-585-1	Sequence 1, App1

ALIGNMENTS

RESULT 1									
US-09-404-879A-72									
; Sequence 72, Application US/09404879A									
; Patent No. 6468546									
; GENERAL INFORMATION:									
; APPLICANT: Mitcham, Jennifer L.									
; APPLICANT: King, Gordon E.									
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND									
; TITLE OF INVENTION: DIAGNOSIS OF OVARIAN CANCER									
; FILE REFERENCE: 210121.462C2									
; CURRENT APPLICATION NUMBER: US/09/404,879A									
; CURRENT FILING DATE: 1999-09-24									
; NUMBER OF SEQ ID NOS: 393									
; SOFTWARE: FastSeq for Windows Version 3.0									
; SEQ ID NO 72									
; LENGTH: 2017									
; TYPE: DNA									
; ORGANISM: Homo sapien									
US-09-404-879A-72									
Query Match									
Best Local Similarity 55.4%; Pred. No. 4.3e-116;									
Matches 1128; Conservative 0; Mismatches 812; Indels 96; Gaps 6;									
QY	226	GGCGTCGATCAGCAAGTGAAGCTATAGAACCATGCTCAGTTCCACACCTTCGGT	285						
DB	9	GGCTGAGAGCTGCAAGAGAGTCAAGATCATGATGCTCAGTTCCACAGCATGAAT	68						
QY	286	GGTAGCCTGATGCTGCGCCATTAAGTGAAGGAGCCAGCATGACGACAGTTC	345						
DB	69	GAGAGGCCAAATATGTGGCTATTAATCTGAAGAAAGTAAAGATGAAGTAAAGTTT	128						
QY	346	CTTAGCCTGAAGCCGATTAAGGATTTATTAAGTGAAGGAGGAGGAGGAGGAGGAGG	405						
DB	129	GATTAACCTTAACCTTCAGAGGATTAATCAATGAGGATGAGGATGAGGATGAGGATG	188						
QY	406	CAATCGGGTTCCTCAGCCTGCTTAGCAAAATATGGCGGTGAGGAGGAGGAGGAGGAGG	465						
DB	189	CAGTCAGCTGCTGCGGCGGCTTTAGCTGAATATGGCCCTTAATCAATCTGAACAAAG	248						
QY	466	GATGAGAGATGATCAATGATGATTTTCCATAGCATGAGGATGATGATGATGATGATGATG	525						
DB	249	GATGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	308						
QY	526	CAGGATATGAGTCTCCCTCCACACTTCCCTGTCATGAAACAGCAACGACGAGT---GCT	582						
DB	309	CAGGAGCAACAGTGTCTGTAGTCTCCCTCCCTATCATCAAAACCCCTATGTTCTCT	368						
QY	583	ATTTCAGGACCAAGCATTTGTATAGAGGAGGATTCATGACATG-----CCACCA	633						


```

Db      369  CCACTAATCTGCTGCTGTTTGGGATGGGAGCATGCCCAATCTGTCATTCATGAGCA 428
Qy      634  CTCACAGCTGTGCTCTGTGCTCA-----TGCGCTCAATT 669
Db      429  TTGCTCTCAGTTGCACTATAGCACACCCTGTCTCTGCTACTTCAGGGGCCGATATT 488
Qy      670  CCAAGTGTGGAAATGCTCCACCCTTAGATCTTGTGCTCCCGACAGAGTGCTGCC 729
Db      489  CTTCCCTTAATGATGCTGCTGCTTACGTGCTTGTAGTACATCTCATTTACCAAT 548
Qy      730  CTGGCTAACGGGCTCTCCCGCTCATAGAGCTGCTGCGTTTGGCATCTCCAGCC 789
Db      549  GGAATGCGCAGTCTCATTCAGCTTTATCCATTCCTTATCTTTCAACATGCTCAT 608
Qy      790  ACATGGCCAAAGATTTCTTCTTACAGATTTGGTCCAGGCTACAAATTAAACTTAG 849
Db      609  GCATCATTTTACGCTGATGATGGAGGATTTGGTGGTGTATGATCCAGAACGCCAG 668
Qy      850  TTACGAAAGGCAATCATGATGATGCGCAGCGCCCTCCAG-----892
Db      669  TCTCTGATTTGATTTGGAITGATGCTCAACTTCTCACTGCTTCCCTCTCAGGGAC 728
Qy      893  -----CAGCAGAAATGGGCTGTGCTCAGTCAATCAAGGCTGAATATACAG 936
Db      729  TCACCTAAGACAGGAGCTCAGAGTGGGAGTTCTCAGCCTTCAAGATTAAAGTATCG 788
Qy      937  CAGTTATTCAACAGCCACGACAAATATGATGAGCACTTACAGTCTCCAGGCAAGA 996
Db      789  CAAAATTTAATATGCTAGACAAAGCAGATGAGCGATACCTCTCAGTTTTCAGACTGA 848
Qy      997  ACTATTTCATGCAATCAAGTTTACCAGGCTCAGGCTCAATATGATGATTTCT 1056
Db      849  AATGCCCTTCTTCACTCAATCTCTCTCAACTCAGTACTACTATTGACTTGCT 908
Qy      1057  GACATTTGATCAGATGAGAAATCTCAGCAGAAATTTATCTTACTATGACCTTAAT 1116
Db      909  GACATCGATGATGAGAGAGAGTGAAGCTGAAGATTTATTTCTGGCATGACCTCACT 968
Qy      1117  GATGTTGCCATGTCTGTGACGCACTGCCCGCTCTGCTCAGAAATCATCTCTCT 1176
Db      969  GACATGGCCAAAGCTGAGCAGCCACTACACTGACGTTGCCCTCCGACCTGCTCTCA 1028
Qy      1177  TCTCTCAGAAGGTTCCGCTCCGGCAGTGGAAATGTCCTGATTAAGTCTTCTTCTGAT 1236
Db      1029  TCTTTCAAGG-----GGAAAGCAAGTTGATTTCTGTAAT 1064
Qy      1237  CAGAGGCTGCTGAGAGACCGTCTCAGAGATGAGCAGCAGCCAGAGAAATCTGCT 1296
Db      1065  GGAATCTGCTTCAATCAAGAAACACAGAGAG---AAGAGCTCAGAAAGAACTGCCA 1121
Qy      1297  GTGACATTTGAAGATTAAGAGCGGAGAACTTCCAGCAGAGCAGTGTGAGTGGAGAAG 1356
Db      1122  GTTACTTTTGGAGCAAAACGGAAGCCAACTATGAAACAGAGAAACATGAGTGGAGAAG 1181
Qy      1357  CGCGCCCAAGGCTCTTGGAGAGAGCGCAAAAGCAGAGCGGTTGGCTCAGCTGAG 1416
Db      1182  CGACGCCAAGTGTGATGAGAGAGCAGAGGAGGCTGAAACGAAAGCCAGAAAG 1241
Qy      1417  CGCGCCGAGCAGAGAGAGAGAGCGGAGCGCCAGAGCAGAGAGCCAAAGCGAGCTG 1476
Db      1242  AAGGAAGATGGAGAGCGAAACAGAGAACTGCAAGAGCAGAAATGGAAGAGAGCTG 1301
Qy      1477  GAGCTGAGAAAGCAGCTGAGAGAGCGGAGCTGAGCGGACGAGAGAGAGAGAG 1536
Db      1302  GAGTTGAGAAACGCTTGGAGAAACAGAGAGCTGAGAGACAGCGGAGAGAGAG 1361
Qy      1537  AGGAAGAGATCGAAGAGCGGAGCGCCAAACCGGAACTGGAAGAGCAGAGCAATT 1596
Db      1362  AGAAAGAGATGAAAGAGCAGAGGCGCAAAACAGAGCTTGAAGACAAAGCGCTTGA 1421
Qy      1597  GAATGGAAACGGAACCGAGAGACAGAACTCTGATCAGAGGAACAAGAGAGAGGCG 1656

```

```

Db      1422  GAATGGAAAGACTCCGTCGACAGAGCTGCTCAGTCAGAAAGACCGAGGAACAAGAGAC 1481
Qy      1657  ACCGTGCTCTGAAGGAGAGAGAACTCTGAGATTGATGATGAAGCTCTGAATAC 1716
Db      1482  ATTGTACAGCTGACCTCCAGAAAGAAAGTCTTCCACCTGGAACCTGGAAGCAGTAATGA 1541
Qy      1717  AAAAGCATCAGCTAGAGAGAAACTTACAGATATCAGAGTGTGCTGCTGCGCAACCAGAG 1776
Db      1542  AACATCAGCAGATCTCAGGAGACTACAGATGTCCAAATCAGAAAGCAACAAAG 1601
Qy      1777  CAAGAAATTGAGACAGAACAGTCTAGAGAGCTAAGAAATTTGTAATATCAACCACTTA 1836
Db      1602  ACTAGCTAGAAATTTTGAATTAACAGTGTGACCTGGAATTTAGAAATCAACCACTT 1661
Qy      1837  CAGCAGAGTTGACAGAAATCTCAGCAATGCTTGGAAAGCTTATTCAGAAACAGATA 1896
Db      1662  CAACAGAGCTTAAAGAAATCAAAATTAAGTTATCTATCTGGTCCCTGGAACAGCTA 1721
Qy      1897  CTCAGTACCACTTAAACAAAGTCCAGCAGAACGTTTGCATAGAGACTGCTTTPACC 1956
Db      1722  TTTAAACGAAAGATTTAAACATGCTCAGTCACTGTAACACCTGATTCAGGATCAGTTTA 1781
Qy      1957  CTCAAAAGAGCTTGAAGCAAGAGAGCTGCCGAGCAGCTCCGGAGCAGCTGAC 2016
Db      1782  CTTCATTAAGATCATCAGAAAGAAAGATTTATGCCAAAGACTTAAAGAACTTAAGAT 1841
Qy      2017  GAGTGGAGAGAGACCAAGTCAAGCTGAGAGATGATGTTTCAACCAACAGCTG 2076
Db      1842  GCTCTGAAAAGAACTGACATCTTACCTTCAAAATGATTTCAATTAACATCAGCTG 1901
Qy      2077  AAGGATCGAGAGATACATACAGCAACAGCACTCCAGAAAGCAGAGTCCCTGAGGCA 2136
Db      1902  AAGGAATCAGAGAAAGCTATATACAGAGATTTAGCCCTTGAACAACCTTCAATTAATC 1961
Qy      2137  GCGGAGCTGAAGCAGAAAGAGCAGAGAGAGAGCTGAGTTAGAGAGCAAA 2192
Db      1962  AACGTGACAAATGGAAGAAATCGAAAGAAAGATTAGCAAAAAA 2017

```

```

RESULT 2
US-09-338-933-72
; Sequence 72, Application US/09338933
; Patent No. 6488931
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer Lynn
; APPLICANT: King, Gordon E.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY OF
; FILE REFERENCE: 210121.462C1
; CURRENT APPLICATION NUMBER: US/09/338, 933
; NUMBER OF SEQ ID NOS: 312
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 72
; LENGTH: 2017
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-338-933-72

```

```

Query Match      9.5%; Score 484.8; DB 4; Length 2017;
Best Local Similarity 55.4%; Pred. No. 4.3e-116;
Matches 1128; Conservative 0; Mismatches 812; Indels 96; Gaps 6;

Qy      226  GGCCTGATCAGCAAGGTGAACGTAATGAAACATGCTCAGTTTCCACACCTTTGGT 285
Db      9    GGTGAGAGCTGCAAGAAAGTCAAGATCATGATGCTCAGTTTCCACAGCATGAT 68
Qy      286  GGTAGCTGATGCTGCGCCATTACTGTGAGAAAGCGCCAGCATGACCAAGCAAGTTC 345
Db      69  GGAAGGCCAAATATGATGGCTATTACATCTGAAGAGGTACTTAAGCATGATTAACAGT 128
Qy      346  CTTAGCTGAAGCGGATAGCGGATTAATTAATCTGATCAAGCGAGAACTTTTTC 405

```

Db 129 GATTAACCTCAACCTTCAGAGGTTACATAACAGTGATCAAGCCCTGCTTTTCTTA 188
 QY 406 CAATCTGGTTTACCTGACCTGTTCTTAGCAATAATATGGGGCTAGGGGACATGAATAC 465
 Db 189 CACTCAGGTTGCGGGCCCGGTTTATAGCTGAATATGGGCTTATCAGATCGAAGAC 248
 QY 466 GATGGAAGATGATCAAGTGAATTTTCCATAGCCATGAAGCTTATCAATGAAGCTA 525
 Db 249 GATGGGAAGATGACGACGAAGGTTCTATAGCTATGAATCAATCAAGTTAAAGTTG 308
 QY 526 CAAGAATATAGCTCCCTCCCAACTTCCCTGTCATGAACAGCAACCAAGT--GCT 582
 Db 309 CAGGGCCCAACGCTGCTGATGCTCCCTCTCATATGAAACAACCCCTGATGTTCT 368
 QY 583 ATTTCAGTGCACACAGATTGGATATGAGAGGATTTGCTAGCTG-----CCACA 633
 Db 369 CCACTAATCTGCTGCTGTTGGATGGGAAGATGCCAATCTGTCATTCATCAGCCA 428
 QY 634 CTCACAGCTGTTGCTCTGTCGCA-----TGGGCTCAT 669
 Db 429 TTGCTCCAGTTGACCTTATAGCAACCTTGTCTTCTGCTACTTCAAGGACACGATTT 488
 QY 670 CCAAGTTGTTGAATGTCCTCACTTATGATTTCTGTCCTCAGACAGAGTGCCTCC 729
 Db 489 CTTCCCTAATGATGCTGCTGCTCCCTAGTCTTCTGTTAGTACATCTCATTAACCAAT 548
 QY 730 CTGGCTAAGCGGGCTCCTCCCTCATACACCTCTGCTGCTGCTTGGGATCCTGCAACC 789
 Db 549 GGAATGCAAGTCTCATTTCACTTATTCATTCCTTATTTCTTCAACATTTGCTCAT 608
 QY 790 ACATGAGCCAAAGATTCTTCTTCAAGAGATCTGTCAGAGGTCAACAATTAACAATAAG 849
 Db 609 GATATCTTTAAGGCTGATGATGAGAGATTTGGTGTGCTGATGATCCAGAAAGCCGAG 668
 QY 850 TTAAGAGGCAATATCATTCATGTCGACGCGCCCTCAG-----892
 Db 669 TCTCTATGATTAGATTAAGATTAAGTCAACTTCCCACTGCTTCCCTCAGGGAAC 728
 QY 893 -----CAGAGAAATGGGCTGTCCTTCACTGATCAACAGGCTGAATACAG 936
 Db 729 TCACTTAAGACAGGACCTTGAAGTGGGAGTTCCTCAGGCTTCAAGATTAAGATTCGG 788
 QY 937 CAGTTATTAACACGACGACGAACTATGATGAGCACTTAACAGGTTCCCGACGACGA 996
 Db 789 CAAAATTTAATGTTAGTGAAGGATGAGGAGTACTCTCAGGTTTCAAGCTTGA 848
 QY 997 ACTATTTCTATGATCAAGTTTAACTCCAGGCTCAGTGGCTTCAATATGAATCTTCT 1056
 Db 849 AATGCCCTTCTTCACTCAATCTCTCAAACTCAGTAGCTACTATTGGACTTGGCT 988
 QY 1057 GACATTTGATCAAGATGGAATCTCATCTGAGAAATTTTCTAGCTATGACCTTAAT 1116
 Db 909 GACATGATGATGACGACGAGTGAAGCTGAAGAAATTTATCTGGGATGACCTCAGCT 968
 QY 1117 GATGTTGCAATGCTGTCAGCAGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCT 1176
 Db 969 GACATGCGCAAGCTGACGACGACCTAATCACTGCTGCTGCTGCTGCTGCTGCT 1028
 QY 1177 TCTTTGAGAGGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1236
 Db 1029 TCTTTGAGAGG-----GGGAAAGCAAGTTGTTCTGTTAAT 1064
 QY 1237 CAGAGGCTGCTGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1296
 Db 1065 GGAATCTGCTTCAATATCAAGAAACACAAAG---AAGAGCTCAGAGAAATCTGCCA 1121
 QY 1297 GTGACATTTGAAGATTAAGAGCGGAGAACTTGAAGAGGAGCTGAGGAGCTGAGAG 1356
 Db 1122 GTTACTTTTGAAGACAAACGGAAGCCCACTATGAACGAGAAACATGAGCTGAGAGAG 1181
 QY 1357 CGCGCCAAAGCGCTCTTGAAGACGACGCGAAAGACAGAGAGGCTGCTGAGCTGAG 1416
 Db 1182 CGACGCCAAGTTGATGAGGACGACGAGAGGAGGCTGAAACGCAAGGCCAGAAAGAG 1241

QY 1417 CGCGCCGAGAGAGAGAGAAAGCGGAGCCGACAGAGACAGAGGCGCAAGCGCAGCTG 1476
 Db 1242 AAGGAAGATGGAGGAGGAGAAACAGAGAACTGCAAGAGCAAAATGAGAAAGCAGCTG 1301
 QY 1477 GAGCTGGAAGAGAGAGCTGAGAGAGCGGAGCTGAGCGGAGGAGAGAGAGAGAG 1536
 Db 1302 GAGTTGAGAGAAAGCTTGGAGAGAAACAGAGAGAGCTGAGAGAGAGAGAGAGAG 1361
 QY 1537 AGGAAGATGAGAGAGCGGAGGCGGAGCGGAGAAACGAGAACTGAGAAAGCAGCAACT 1596
 Db 1362 AGAAAGAGATGAGAAAGAGAGAGGAGGAGCAAGAAACAGAGCTTGAAGACAGCCCTTTA 1421
 QY 1597 GAATGGGAACGGAACCGGAGAGAGAACTTCTGTAATCAAGAGAAACAGAGAGAGAGG 1656
 Db 1422 GAATGGGAAGAGCTTCTGCGAGAGAGCTGCTCAAGAGAGCCAGGAAACAGAGAGAG 1481
 QY 1657 ACCGTGTCCTGAAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1716
 Db 1482 ATTGTAGGCTGAGCTTCAAGAAAGAAAGTCTTCACTGAGAGAGAGAGAGAGAGAG 1541
 QY 1717 AAAAGCATGAGCTGAGAGAGAAACCTTCAAGATATCAGAGTGTGAGCTGAGCAACCAAG 1776
 Db 1542 AAACATCAGAGAGATCTCAGGAGAGCTAAGAGATGTCGAATCAGAAAGCAAGAG 1601
 QY 1777 CAGAAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1836
 Db 1602 ACTGAGCTGAGAGATTTGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1661
 QY 1837 CAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1896
 Db 1662 CAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1721
 QY 1897 CTCAGTACAGTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1956
 Db 1722 TTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1781
 QY 1957 CTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2016
 Db 1782 CTTGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1841
 QY 2017 GAGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2076
 Db 1842 GCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1901
 QY 2077 AAGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2136
 Db 1902 AAGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1961
 QY 2137 GCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2192
 Db 1962 AAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2017

RESULT 3
 US-09-215-681-72
 ; Sequence 72, Application US/09215681A
 ; Patent No. 6528253
 ; GENERAL INFORMATION:
 ; APPLICANT: Mitcham, Jennifer L.
 ; APPLICANT: Prudakis, Tony N.
 ; APPLICANT: King, Gordon B.
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSIS
 ; FILE REFERENCE: 210121.463
 ; CURRENT APPLICATION NUMBER: US/09/215,681A
 ; NUMBER OF SEQ ID NOS: 310
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 72
 ; LENGTH: 2017
 ; TYPE: DNA
 ; ORGANISM: Homo sapien

US-09-215-681-72

Query Match 9.5%; Score 484.8; DB 4; Length 1017;
Beet Local Similarity 55.4%; Pred. No. 4.3e-116;
Matches 1128; Conservative 0; Mismatches 812; Indels 96; Gaps 6;

226 GGCCTGATTCAGCAAGTGAAGTAAATAGAACATGAGTCTCAGTTTCCACACCTTCCGT 285
Db GGCCTGATTCAGCAAGTGAAGTAAATAGAACATGAGTCTCAGTTTCCACACCTTCCGT 285
9 GGCCTGATTCAGCAAGTGAAGTAAATAGAACATGAGTCTCAGTTTCCACACCTTCCGT 68
286 GGTAGCTGATGTCGTGGCCATTAATGTAAGAAAGGCGCAGCATGACAGAGTTTC 345
Db GGTAGCTGATGTCGTGGCCATTAATGTAAGAAAGGCGCAGCATGACAGAGTTTC 345
69 GAGAGGCCAAATATATGCGCTATTACATCTGAAGAACCTAAGCATGATATAACAGTTT 128
346 CTTAGCTGAAAGCCGATAGCGGATTTATTAATGTAATCAAGCAGAGAACTTTTTC 405
Db CTTAGCTGAAAGCCGATAGCGGATTTATTAATGTAATCAAGCAGAGAACTTTTTC 405
129 GATTAACCTCAAACTTTCAGAGGTTACATACAGGTATCAAGCCGTAATTTTCTTA 188
406 CAATCTGGGTTACCTCAGCCTGTCTTAGCACAAATATAGGCGCTAGCGAGCATGAAATAC 465
Db CAATCTGGGTTACCTCAGCCTGTCTTAGCACAAATATAGGCGCTAGCGAGCATGAAATAC 465
189 CAGTCAGGTCTGCGGCGCCGTTTACGTAAATATGCGCTTATCAGATCTGAACAG 248
466 GATGGAAGATGATCAAGTGAATTTTCAATAGCCATGAAGCTTATCAAACTGAAGCTA 525
Db GATGGAAGATGATCAAGTGAATTTTCAATAGCCATGAAGCTTATCAAACTGAAGCTA 525
249 GATGGAAGATGATCAAGTGAATTTTCAATAGCCATGAAGCTTATCAAACTGAAGCTA 308
526 CAAGATATTCAGCTCCCTCCACACTTCCCTGTGATGAAACAGCAACAGTG---GCT 582
Db CAAGATATTCAGCTCCCTCCACACTTCCCTGTGATGAAACAGCAACAGTG---GCT 582
309 CAGGCGCAACAGCTGCTGATGCTCCCTCATCATGAAACAAACCCCTATGTTCTCT 368
583 ATTTCAGTGCACAGCAATTTGATAGAGGATTTGATGATG---CCACCA 633
Db ATTTCAGTGCACAGCAATTTGATAGAGGATTTGATGATG---CCACCA 633
369 CCATCAATCTGCTGCTTTTGGATGGAAAGCATGCCCAATCTGCAATTCATAGCA 428
634 CTCACAGCTGTTGCTCTGTGCA---TGGGCTCAAT 669
Db CTCACAGCTGTTGCTCTGTGCA---TGGGCTCAAT 669
429 TTGCTCCAGTTGCACTATAGCAACACCTGTCTTCTGCTACTTCAAGGACAGTAT 488
670 CCAATTTTGAATGTCACACCTTATGATTTTGTGCTCCAGCAGAGTGTCTCC 729
Db CCAATTTTGAATGTCACACCTTATGATTTTGTGCTCCAGCAGAGTGTCTCC 729
489 CTTCCCTTAATGATGCTGCTCCCTAGGCTTCTGTAATGATCTCATTAACAAAT 548
720 CTGGTAAAGGGGCTCTCCGCTCAATACAGCTCTGCTGCTTGGCATCTGACAC 789
Db CTGGTAAAGGGGCTCTCCGCTCAATACAGCTCTGCTGCTTGGCATCTGACAC 789
549 GGAATCTGCAATCTATTCAGCTTTATTCATCTTATTTCTTCAACATGCTCAT 608
790 ACATGCGCAAGAGTTCTTCTTCAAGATCTGCTCAAGGCTCAAACTAAACCTAAG 849
Db ACATGCGCAAGAGTTCTTCTTCAAGATCTGCTCAAGGCTCAAACTAAACCTAAG 849
609 GCATCATCTTACAGCTGATGATGGAGATTTGGTGGCTAATATCAGAAAGCCAG 668
850 TTACGAAGGCAATCATGATGTCGCGCAGCGCCCTCCAG----- 892
Db TTACGAAGGCAATCATGATGTCGCGCAGCGCCCTCCAG----- 892
669 TCTGTATTTGATTTAGATCTAGTACCTCACTTCTCAACGCTTCCCTCAGGGAAC 728
893 -----CAGCAGATGAGGCTGTGCTCAGTCAATCAAGGCTGAATATACAG 936
Db -----CAGCAGATGAGGCTGTGCTCAGTCAATCAAGGCTGAATATACAG 936
729 TCACCTAAGACAGGAGCTCAGAGTGGGAGTTCTCAGCTTCAAGATTAAGTATGG 788
937 CAGTTATTCACAGCAGCAAACTATAGTGAACCTTAAACAGTCTCCAGGCAAGA 996
Db CAGTTATTCACAGCAGCAAACTATAGTGAACCTTAAACAGTCTCCAGGCAAGA 996
789 CAAAATTTAATATAGCTAGACAAAGCATGACGATACCTCAGGTTTCAACCTGA 848
997 ACTATTTCAATCAATCAAGTTTACCCAGGCTCAGTGGCTTCAATATGAATTTTCT 1056
Db ACTATTTCAATCAATCAAGTTTACCCAGGCTCAGTGGCTTCAATATGAATTTTCT 1056
849 AATGCTCTTCTTCAATCAATCTCTCAACCTCAGCTAGCTAATTTTGGCTGTGCT 908
1057 GACATTTGATCAAGATGAGAACTCACTGAGAAAGATTTATCTGATAGCACTAAT 1116
Db GACATTTGATGATGAGAACTCACTGAGAAAGATTTATCTGCGATGACCTTCACT 968
909 GACATGATGATGAGAACTGAGAAAGTGAAGATTTATCTGCGATGACCTTCACT 968
1117 GATGTTGCGATGTCGTGACGCACTGCGCGCTGCTGCTCAGAAATACATCCCTCT 1176
Db GATGTTGCGATGTCGTGACGCACTGCGCGCTGCTGCTCAGAAATACATCCCTCT 1176
969 GACATGCGCAAGCTGAGCAAGCACTACAGCTGAGGTTGCTCCGAGGTTGTCTCTCA 1028

QY 1177 TCCTTCAAGAAAGTTCGCTCCGCGATGGATGTCCTCATTAAGCTCTTCTGTGAT 1236
Db 1029 TCTTTCAGAG-----GGAAAAGCAAGTTGATTTCTGTTAAT 1064
QY 1237 CAGAGCTGCTGAGAGCCGCTGTCAGAGATGAGCAGCAGCCAGAAAGAACTGCT 1296
Db 1065 GGAATCTGCTTCTATTCAGAAAACACAAAGAG---AAGGCTCAGAAAGAACTGCA 1121
QY 1297 GTGACATTTGAATTAAGAAAGCGGAGAACTTCAAGCAGGCACTGTGAGCTGAAG 1356
Db 1122 GTTACTTTTGAAGCAAAACGAAAGCCAACTATGAAAGAGAAACATGAGCTGGAAG 1181
QY 1357 GCGCCGAAAGCTCTTGAAGCAGCAGCGCAAGAGAGAGCGGCTTGGCTCAGCTGAG 1416
Db 1182 CGACCCAAAGTTGATGAGACAGCAGCAGAGAGGCTTAAGCCAAAGCCAGAAAGAG 1241
QY 1417 GCGCCGAGCAGAGAGAGAAAGAGCGGAGCGCCAGAGCAGAGAGCCAGCGCAGCTG 1476
Db 1242 AAGGAAGAGTGGAGCGGAAACAGAGAACTGCAAGAGCAAGATGAAAGAGCAGCTG 1301
QY 1477 GAGCTGAGAAAGCAGCTGAGAGCAGCGGAGCTGAGAGCGGCGAGAGAGAGAG 1356
Db 1302 GAGTTGAGAAAGCCTTGGAGAAACAGAGAGAGCTGAGAGACAGCGGAGAGAGAGG 1361
QY 1537 AGGAAGAGATCGAGAGCGGAGCGGCAAAACGGAGCTGAAAGCAGCGCACTT 1596
Db 1362 AGAAGAGATGAGAAAGAGAGAGAGCAGAAAGAGAGCTTGAAGACAAAGCGCTTTA 1421
QY 1597 GAATGGAAGCAGAACCGGAGCAGAACTCCTGATCAGAGAGCAAGAGCAGAGAGG 1656
Db 1422 GAATGGAAGAGCTCCGTCGAGAGAGCTGCTCAGAGAGCAGAGAGCAAGAGAGAG 1481
QY 1657 ACCGTGCTTGAAGCAGAGAGAGAGAGCTTGAAGTTGAGTTAAGAGCTTGAATGAC 1716
Db 1482 ATTGTCAAGCTGAGCTCAGAAAGAGAGAGCTCCAGCTGAACTGAAAGCAGTAAATGGA 1541
QY 1717 AAAAGATCAGCTAAGAAAGAACTCAGAGATTCAGTGTGCTGCTGAGCAACCCAGAG 1776
Db 1542 AATCATCAGAGATCTCAGGAGACTCAGAGATTCAGAAATCAGAAAGCAAAAG 1601
QY 1777 CAGAAATGAGAGCAGAACTCAGAGATTCAGAGAGTAAAGTGTGTAATCAGCCACTTA 1836
Db 1602 ACTGAGCTAAGAAAGTTTGAATTAACAGTGTGAGCTGAAATTAAGAAATCAACACTT 1661
QY 1837 CAGAGCAGTTGAGAGATCTCAGCAATGCTTGAAGACTTATTCAGAGAAACAGATA 1896
Db 1662 CAGCAGAGCTTAAAGAAATCAAAATAGCTTATCTATCTGCTCCCTGAGAAAGCAGCTA 1721
QY 1897 CTCAGTACCAATTAACAAAGTCCAGAGACAGTTTGCATAGAGCTGCTTACC 1956
Db 1722 TTAAACGAAGATTTAAACATCAGCTCAGTACACACTGATTCAGGATAGTTTA 1781
QY 1957 CTCAAAAGAGCTTGAAGCAGAAAGAGCTGCGCGGAGCAGCTCCGAGAGCAGCTGAGC 2016
Db 1782 CTTCAATAAAGTATCAGAAAGAGAAATTAAGCAAGACTTAAAGAACATTAATGAT 1841
QY 2017 GAGGTGAGAGAGAGCAGGCTCAAGCTGAGAGAGTTGATTTTCAACACAGCTG 2076
Db 1842 GCTCTTGAAGAAAGAACTGATCTAGCTCTCAAAATGATGATTAATTAACATCAGCTG 1901
QY 2077 AAGGATGAGAGAGATTAATAGCAACAGCACTCCAGAGCAGAGGCTCCGAGAGCA 2136
Db 1902 AAGGATCTCAGAGAAAGCTTAATTAACAGAGCTTAGCCCTTGAACAACTTCAATTAATC 1961
QY 2137 GCGGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2192
Db 1962 AAGCTGACAAATTTGAAGAGAAATGGAAGAAAGATTAAGGCAAAAAAAAAAAAA 2017

RESULT 4
US-08-630-915A-193
; Sequence 193, Application US/08630915A

```

: Patent No.6309820
: GENERAL INFORMATION:
: APPLICANT: SPARKS, Andrew B.
: APPLICANT: HOFFMAN, No. 6309820h
: APPLICANT: RAY, Brian K.
: APPLICANT: FOWKES, Dana M.
: APPLICANT: MCCONNELL, Stephen J.
: TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
: TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
: NUMBER OF SEQUENCES: 227
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Pennie & Edmonds LLP
: STREET: 1155 Avenue of the Americas
: CITY: New York
: STATE: New York
: COUNTRY: USA
: ZIP: 10036-2711
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/630,915A
: FILING DATE: 03-APR-1996
: CLASSIFICATION: 536
: ATTORNEY/AGENT INFORMATION:
: NAME: Mistrock, S. Leslie
: REGISTRATION NUMBER: 18,872
: REFERENCE/DOCKET NUMBER: 1101-174
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 790-9090
: TELEFAX: (212) 869-8864/9741
: TELEX: 66141 PENNIE
: INFORMATION FOR SEQ ID NO: 193:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2873 bases
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: unknown
: MOLECULE TYPE: DNA
: US-08-630-915A-193

Query Match          6.6% ; Score 333.2 ; DB 4 ; Length 2873 ;
Best Local Similarity 61.0% ; Pred. No. 1.7e-76 ;
Matches 633 ; Conservative 0 ; Mismatches 363 ; Indels 42 ; Gaps 4 ;

OY 2867 CTTCTCTACCCGTACTAGTGTGTGGCCAGGTACGCGACAGATCAGCCTTTACCCAGCGCA 29286
Db 541 CTTTTCCTAACCTAATCTTAATATCATCATGCGAGAAAAATCAAGCCTTCACTCGAAGCTG 600
OY 2927 CAGCCACTGGCGCTCCCATCTCCCGTCCCTGGGCGCAGGGTGAAGAAAGGTGAAGGGCTAC 29866
Db 601 TGTCCCTCG--ATCTGTATCACTTAATTCATGACAGAGGACAGATGTGAAGAACTTAA 657
OY 2987 AAGCGCAAGCCCTGTATCCCTGTGAGAGGCGCAAAAAAGACAAACACTTAATTTTAACTAAA 30468
Db 658 AAGCACAAGCCCTTTGTTCTGCTGAGATCGAAGAAAGAAATTAACACCTGAACCTTCATAAAC 717
OY 3047 GTGACGTATACCGTTCTTGTGGAACGACGACAGACATGTGTGTTGTGGAGAGTTCAAGTCT 31068
Db 718 ATGACATTAATTAATCTGTCTTGGAGCGCAAGAAATTTGTGTGGTGGGAGGTGATGAG 777
OY 3107 AGAAGGGTTGGTGTCCCAAGTCTTAAGTGAAGTCAATTTCAAGGGCCGTAAGGAATCCA 31666
Db 778 GAAAGGATGTGTTTCCCAATCTTAATGTCAAGATATATTCCTGGAGGTAAAGTAAACGCG 837
OY 3167 CAAGCATGATACTGGCCCTACTGAAAGTCTGTACTAGTCTTAAAGAGAGTGGCTTCCCGG 32266
Db 838 AA-----GAAACAGAGCTTTGTATGACGCTGTAAATTAAGAAACTACCTGCG 885
OY 3227 CCGCCAGACCGCACTTCCCGGAGAGAGTTATTATTCATCATGTACATACAGAGTTCTG 32866

```

```

Db      886  CAGCCTTATTCAGT-----TGAGAGAAATATATGTGCACTTTATCCATATTCAGAGTG 939
Qy      3287 AGCAAGAGATTTTACCTTTGAGCAGAGGGATGTGATTTGTGTATCCAGAAAGATGTG 3346
Db      940  AACTGGAGATTTGACTTTGACAGAGAGTGAAGAAATATTTGTGTGACCCAGAAAGATGTG 999
Qy      3347 ACTGTGTGACGGGACCGGTGTGGGCGACAGATGCCGAGTCTTCCCTTCTAATATGTGAGGC 3406
Db      1000 AGTGTGTGACAGGAGATTTGTGAGATGAGATGTGAATTTTCCATCAATATGTCAAA 1059
Qy      3407 TTAAGAATTCAGAGGGCTGTGAACTGTGTGGAAAACAGGAGATTTAGAAAAAACCTG 3466
Db      1060 CAAAGGATCAAGAGATTTTGGAGGTGTAGCAAGTCTGAGGATCAATTAATTAACCTG 1119
Qy      3467 AATTTGCCAGAGTATTTGCTTCCATCGCTCTACTGTGTCCGAAACAATCACTGCTGCTC 3526
Db      1120 AGATTTGCTCAGATTAATCTTCAAGCATATGTGTCTTGTGTGTGAAACAATTAAGCTTGAC 1179
Qy      3527 CTGGGCAAGCTGATTTCTGATCCGAAAAAAGAACCCAGGTGATGTGTGGAAAGAGAACCTG 3586
Db      1180 CAGGACAGTTAAATATTTAATCTTAAGAAATAATCAAGTGTGTGTGGCAAGAGATTAC 1239
Qy      3587 AAGCTCGAGGAAAAAACCCGCAATAGGTGTGTTCAGCAATTAATGTCAAATCTTAA 3646
Db      1240 AGGCGCAAGGAAAAAAGCCGACAGAAAGATGTGTTCCTGCAAGTCAAGTTAACTTTTG 1299
Qy      3647 GCCCCGGAACAAGCAAAATACCCCAACTGAGTACCAGAGACCGAGTGCAGCCAGCAG 3706
Db      1300 GTCCAAATGATGAAAGAGCCACACTGTCTTATC-----CTG 1338
Qy      3707 TGTGCGAAGTGTATGGGATGTAGATTAACAACCGCCAGAACATGACAACTTCA 3766
Db      1339 TATGTCAAGTATGTGATGTATGATGACATATCAGAGAAATATGAAATGATGAGCTGATTT 1398
Qy      3767 GCAAGGCCAGATATATAAGTCTCTCAAGAGAGAACCCGACCTGTGTGAAAGAGAAC 3826
Db      1399 CCAAGGACAACTTATTAATGTTATGAAACAAAGATGATCTGTGTTGTGCAAGAGAGA 1458
Qy      3827 TCAGTGGGCAAGTTGGGCTCTTCCCATCCAAATTATGTAAGCTGACACAGACATGAGCC 3886
Db      1459 TCACGGGGATGCTGGTCTCTTTTCTTCAACATACGTTAAGATGACAGACACTGATC 1518
Qy      3887 CCAGCCAGCAATGATCA 3904
Db      1519 CAAGTCAACAGTGACCA 1536

RESULT 5
US-08-630-915A-39
/ Sequence 39, Application US/08630915A
/ Patent No. 6309820
/ GENERAL INFORMATION:
/ APPLICANT: SPARKS, Andrew B.
/ APPLICANT: HOFFMAN, No. 6309820h
/ APPLICANT: KAY, Brian K.
/ APPLICANT: FOWLKES, Dana M.
/ APPLICANT: MCCONNELL, Stephen J.
/ TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
/ TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
/ NUMBER OF SEQUENCES: 227
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Pennie & Edmonds LLP
/ STREET: 1155 Avenue of the Americas
/ CITY: New York
/ STATE: New York
/ COUNTRY: USA
/ ZIP: 10036-2711
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS

```

```

SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/630,915A
FILING DATE: 03-APR-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Miarock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-174
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 39:
SEQUENCE CHARACTERISTICS:
LENGTH: 747 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
US-08-630-915A-39

```

```

Query Match      5.1%; Score 258; DB 4; Length 747;
Best Local Similarity 63.9%; Pred. No. 3e-57;
Matches 418; Conservative 0; Mismatches 215; Indels 21; Gaps 1;

QY 3247 GGAGAGAGATTATTCATGTACATACAGAGAGTTCTGAGCAAGAGATTAACTTT 3306
DB 115 GGAGAGAGATATATGCACTTTATCATATTCAGAGTGGAAACCGAGATTGCTTTC 174

QY 3307 CAGCAAGGAGATGATGTGTGTGTACCAAGAAAGTGTGACTGTGTGACCGGAACGGTG 3366
DB 175 ACAGAGAGTGAAGAAATATGTGTGACCAAGAAAGTGAAGTGTGTGACAGAAATAT 234

QY 3367 GCGCAGATCCGAGAGTCTCCCTTCTACTATGTGAGGCTTAAAGATTGAGGCTCT 3426
DB 235 GGAATATAGATGGAATTTTTCATCAATATGTCAAAACCAAGATCAAGAGATTTT 294

QY 3427 GGAATGCTGGAAACAGGAGATTAGGAAAAAACTGAAATTTGCCAGATTATGCT 3486
DB 295 GGGAGTGTACAGAGTGTGAGCATCAATATAAAACCTGAGATTGCTCAGATTAATTCA 354

QY 3487 TCCATCGTGTCTACGTGTCCGGAAGACTCAACCCGTGGCTCTGGGCAAGTATCTGATC 3546
DB 355 GCATATGTTGCTTCTGTGTTCTGAACACTTAGCTTGACCAAGCAAGTAAATTAAT 414

QY 3547 CGGAAAAAGAACCCAGGTGATGTGTGGAAGAGAACTGCAAGCTCGAGGGAAAAAGCGC 3606
DB 415 CTAAGAGAAATACAGATGTGTGTGTGCAAGAGAGTTACAGGCCAAGGAAAAAGCGA 474

QY 3607 CAGATAGGTGTGTTTCCAGCAATATATGTCAAACTTTCTAAGCCCGGAACAAAGCAAAATC 3666
DB 475 CAGAAAGATGTGTTCTGCGCAGTATGTAACTTTGGGTCCAAAGCATGGAAGAACCC 534

QY 3667 ACCCAACTGAGTACCCAGAGCCGAGTGCAGAGCTGTGCAAGTATGCGGAG 3726
DB 535 AACCTGCTTTTCTATCTT-----GTAATGTCAAGTGTCTATG 573

QY 3727 TACGATACACCGCCAGAACGATGACGAACTAGCTTCAGCAAGAGCCAGATCATCAAC 3786
DB 574 TATGCTATGACGAAATATATAGATGAGCTCAATTTCTCCAAAGGAACTCATTAAT 633

QY 3787 GTCTCTCAACAGAGAGACCCGAGCTGTGTGAAAGAGAAAGATCAGTGGCAAGTTGGCTC 3846
DB 634 GTTATGAACAAAGATGATCTGTATGTGTGTGCAAGAGAGATCAAGGGGTGACTGTCTC 693

QY 3847 TTCCTCATTAATTAAGTGAAGTGCACAGACATGAGACCCAGCAGCAAGTAA 3900
DB 694 TTTCTTCAAACTAGTTAAGATGACGACAGCTCAGATCCAAAGTCAACAGTGA 747

```

RESULT 6
US-09-404-879A-5

```

Sequence 5, Application US/09404879A
Patent No. 648546
GENERAL INFORMATION:
APPLICANT: Mitcham, Jennifer L.
APPLICANT: King, Gordon E.
APPLICANT: Algate, Paul A.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
TITLE OF INVENTION: DIAGNOSIS OF OVARIAN CANCER
FILE REFERENCE: 210121.462C2
CURRENT APPLICATION NUMBER: US/09/404,879A
CURRENT FILING DATE: 1999-09-24
NUMBER OF SEQ ID NOS: 393
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 5
LENGTH: 531
TYPE: DNA
ORGANISM: Homo sapien
US-09-404-879A-5

```

```

Query Match      3.0%; Score 151.4; DB 4; Length 531;
Best Local Similarity 61.3%; Pred. No. 1.5e-29;
Matches 264; Conservative 0; Mismatches 161; Indels 6; Gaps 1;

QY 226 GGCCTCGATCAGCAAGGTGAACGTATAGAACCATGCTCAGTTTCCACACCTTTGGT 285
DB 9 GGTGAGAGCTGCAAGAGAAAGTACAGATCATGTGCTCAGTTTCCACAGCATGAT 68

QY 286 GGTAGCTGTATGTCTGGCCATPACTGTGTGAGAAAGGCCCAAGCATGACAGACTTC 345
DB 69 GAGAGGCCAAATATGTGGCTATTTACTATCTGAAGAACGTACTAGCATGTAAACAGTTT 128

QY 346 CTTAGCTGAAGCCGATAGGGGATTTATTTACTGTGTATCAGTCAAGGAGAACTTTTTC 405
DB 129 GATACCTCAAACTTTGAGAGTTACATTAACAGGTATCAAGCCGTACTTTTCTTA 188

QY 406 CAATCTGGTTACTCTCACTGTCTTAGCAAAATATGGGGCTAGCGGACATGAATPAC 465
DB 189 CAGTCAGTGTGCGCGCGCGCGTGTAGCTGAATATGGGCTTATCAGATCTGAACAG 248

QY 466 GATGAAAGATGATCAAGTGAATTTTCCATPACCTAGCAAGTATCAAACTGAAGCTA 525
DB 249 GATGGGAAGATGAGACCAAGAGTCTCTATACCTTGAAGATCAATCAAGTTAAAGTTG 308

QY 526 CAAGATATACGCTCCCTCCACACTTCCCTGTGATGAAGCAAGCAAGTGGCTATT 585
DB 309 CAGGCCAAGCTGCTGTATGCTCTCTCTCTATCAAGAAACACCC-----CTATG 362

QY 586 TCCAGTCAACAGCATTTGTATAGAGGATTTCTAGCATGCGACATCTACAGCTGT 645
DB 363 TTTCTCTCACTAATCTGTGCTGTTTGGATGGAGATGCCCAATCTGTCTATTCAT 422

QY 646 GCTCTGTGCC 656
DB 423 CAGCCATTTGCC 433

```

RESULT 7
US-09-338-933-5
Sequence 5, Application US/09338933
Patent No. 648931
GENERAL INFORMATION:
APPLICANT: Mitcham, Jennifer Lynn
APPLICANT: King, Gordon E.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY OF
TITLE OF INVENTION: OVARIAN CANCER
FILE REFERENCE: 210121.462C1
CURRENT APPLICATION NUMBER: US/09/338,933
CURRENT FILING DATE: 1999-06-23
NUMBER OF SEQ ID NOS: 312
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 5
LENGTH: 531
TYPE: DNA

ORGANISM: Homo sapien
US-09-338-933-5

Query Match
Best Local Similarity 61.3%; Pred. No. 1.5e-29;
Matches 264; Conservative 0; Mismatches 161; Indels 6; Gaps 1;

QY 226 GGGCTGATCAGCAAGTGAACGTAAATAGAACCATGAGCTTCCACACCTTTCCGGT 285
DB 9 GGGCTGAGAGCTGCAAGAAAGATCAGATCATATGAGCTTCCACAGCGATGAAT 68
QY 286 GGTAGCCTGATCTGGGGCCATTAAGTGGAGAAAGGCCAAGCATGACAGCAGTTC 345
DB 69 GGGAGGCCAAATATGAGGCTATTATCATCTGAAGAACGTAAAGATGATTAACAGTTT 128
QY 346 CTTAGCCTGAAGCCGATAGGGGATTTATTAATCTGGATGACAGCGAGAACTTTTTCG 405
DB 129 GATTAACCTTCAAACTTCAGAGAGTTACATTAACAGGTATCAAGCCGTTCTTTTCCCTA 188
QY 406 CAATCTGGGTTACCTCAGCCTGTCTTAGCACAATAATGGGCGCTAGCGGCAATGAATAC 465
DB 189 CAGTCAAGTCTGCGGGCCCGGCTTTTACGTAATATGGGCTTATCAAGTCTGAACAG 248
QY 466 GATGAGAGATGATCAAGTGAATTTTCCATAGCCATGAGCTTATCAAACTGAAGCTA 525
DB 249 GATGGAGAGATGAGCAGCAGAAAGTCTCTATAGCTATGAACCTCAAGTTAAAGTTG 308
QY 526 CAAGATATCAGTCCCTCCACACTTCCCTCTGTATGAAACAGCAACAGTGGCTATT 585
DB 309 CAGGGCCAAAGCTGCTGTAGTCTCCCTCTATATGAAACACCCCTTATG 362
QY 586 TCACATGACACGACATTTGGTATAGAGGAGTTGTAGATGACCACTCAAGCTGT 645
DB 363 TTCTCTCCACTAATCTCTGCTGTTTGGATGGAAAGATGCCAATCTGTCATTCA 422
QY 646 GCTCCTGTGCC 656
DB 423 CAGCCATTGCC 433

RESULT 8

US-09-215-681-5
Sequence 5, Application US/09215681A
Patent No. 6528253
GENERAL INFORMATION:
APPLICANT: Mitcham, Jennifer L.
APPLICANT: Frudakis, Tony N.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSIS
FILE REFERENCE: 210121.463
CURRENT APPLICATION NUMBER: US/09/215,681A
CURRENT FILING DATE: 1998-12-17
NUMBER OF SEQ ID NOS: 310
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 5
LENGTH: 531
TYPE: DNA
ORGANISM: Homo sapien
US-09-215-681-5

Query Match
Best Local Similarity 3.0%; Score 151.4; DB 4; Length 531;
Matches 264; Conservative 0; Mismatches 161; Indels 6; Gaps 1;

QY 226 GGGCTGATCAGCAAGTGAACGTAAATAGAACCATGAGCTTCCACACCTTTCCGGT 285
DB 9 GGGCTGAGAGCTGCAAGAAAGATCAGATCATATGAGCTTCCACAGCGATGAAT 68
QY 286 GGTAGCCTGATCTGGGGCCATTAAGTGGAGAAAGGCCAAGCATGACAGCAGTTC 345
DB 69 GGGAGGCCAAATATGAGGCTATTATCATCTGAAGAACGTAAAGATGATTAACAGTTT 128

QY 346 CTTAGCCTGAAGCCGATAGCGGATTTATTAATCTGATGATCAAGCAGGAACTTTTTC 405
DB 129 GATTAACCTTCAAACTTCAGAGAGTTTACATTAACAGGTATCAAGCCGTTATTTTCTTA 188
QY 406 CAATCTGGGTTACCTCAGCCTGTCTTAGCACAATAATGGGCGCTAGCCGACATGAATAC 465
DB 189 CAGTCAAGTCTGCGGGCCCGGTTTACGTAATATGGGCTTATCAAGTCTGAACAG 248
QY 466 GATGAGAGATGATCAAGTGAATTTTCCATAGCCATGAAAGCTTATCAAACTGAAGCTA 525
DB 249 GATGGAGAGATGAGCAGCAGAAAGTCTCTATAGCTATGAAGAACCTCAAGTTAAAGTTG 308
QY 526 CAAGATATCAGTCCCTCCACACTTCCCTGTGATGAAGAAACAGCAACAGTGGCTATT 585
DB 309 CAGGGCCAAAGCTGCTGTAGTCTCCCTCTATATGAAGAAACACCCCTTATG 362
QY 586 TCACATGACACGACATTTGGTATAGAGGAGATGCTAGCATGCCACATCAGCAGCTGT 645
DB 363 TTCTCTCCACTAATCTCTGCTGTTTGGATGGGAAAGCATGCCCCAATCTGTCCATTCA 422
QY 646 GCTCCTGTGCC 656
DB 423 CAGCCATTGCC 433

RESULT 9

US-09-404-879A-60
Sequence 60, Application US/09404879A
Patent No. 6468546
GENERAL INFORMATION:
APPLICANT: Mitcham, Jennifer L.
APPLICANT: Algate, Paul A.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
FILE REFERENCE: 210121.462C2
CURRENT APPLICATION NUMBER: US/09/404,879A
CURRENT FILING DATE: 1999-09-24
NUMBER OF SEQ ID NOS: 393
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 60
LENGTH: 480
TYPE: DNA
ORGANISM: Homo sapien
US-09-404-879A-60

Query Match
Best Local Similarity 62.1%; Pred. No. 1.2e-27;
Matches 247; Conservative 0; Mismatches 145; Indels 6; Gaps 1;

QY 259 ATGGCTCAGTTTCCACACCTTTCGGTGTAGCCTGTGATGCTGGGCCATTAATCTGTGAG 318
DB 14 ATGGCTCAGTTTCCACAGCATGATGAGAGGCCAATATGTTGGCTATTATCAATCTGAA 73
QY 319 GAAAGGGCCAGCATGACACAGATCTCTTACCTGAGAACCCGATAGGGGATTTATTA 378
DB 74 GAAAGTACTAAGCATGATGAACAGTTGATTAACCTCAAACTTCAAGAGGTTTACATACA 133
QY 379 GGTGATCAAGCAGGAACTTTTTCATCTGGGTTTACCTGACCTGCTTATGACACA 438
DB 134 GGTGATCAAGCCCGATCTTTTCTTACAGTCAAGTCTGCGGGCCGATTTTATGCTGAA 193
QY 439 ATATGGGCGCTAGCGGACATGAATPAAGATGAAAGATGATCAAGTGAATTTTCCATA 498
DB 194 ATATGGGCGCTTATCAAGATCTGAACAGATGGAAGATGACACAGCAAGATCTCTATA 253
QY 499 GGCATGAAGCTTATCAAGTGAAGTCAAGGATATAGCTCCCTCCACACTTCCCTCC 558
DB 254 GCTATGAAGTCAATCAAGTTAAAGTTGAGGCGCAAGAGCTGCTGATGCTCTCTCTCT 313
QY 559 GTCATGAAGACAGCAAGATGCTATTTCAGTGCACAGCACTTGTGATAGAGGATTT 618
DB 314 ATCATGAAGAACACCCCTTATGTTCTCTCCACTAATCTCTGCTCTGTTTGGATG 367

REFERENCE/DOCKET NUMBER: 0575/52268/JPM/MSC/SKS
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-278-0400
TELEFAX: 212-391-0525
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3489 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..3489
US-08-728-323A-1

Query Match 2.8%; Score 144; DB 2; Length 3489;
Best Local Similarity 45.1%; Pred. No. 4e-27;
Matches 534; Conservative 0; Mismatches 650; Indels 0; Gaps 0;

1233 GGATCAGAGGCTGCTGAGGAGCGCTGTCAGAGGATGACAGCAGCCAGAGAAACT 1292
1578 GCAGAGGCCACAGCAGAGGAGCCACAGCAGAGGCCACAGCAGAGGAGCCACAGCA 1637
1293 GCCTGTGACATTTGAAGATAAAGCGGAGAACTTTCAGCGAGGAGTGTGAGCTGGA 1352
1638 GCAGAGGCCACAGCAGCGGAGCCACAGCAGCGGAGGCCACAGCAGCGGAGCCACAGCA 1697
1353 GAAGCGCCGCCAAGCGCTCTTGAAGCAGCGCCAAAGACAGAGCGGTTGCTCACT 1412
1698 GCGGAGGCCACAGCAGCGGAGCCACAGCAGCGGAGCCACAGCAGCGGAGCCACAGCA 1757
1413 GGAGCCGCCGCGCAGAGAGGAGAAAGCGGAGCCGCGCAGACAGACAGAGCCAAAGCGCA 1472
1758 GCGGAGGCCACAGCAGCGGAGCCACAGCAGAGATGACAGCAGAGATGAGAGCA 1817
1473 GCTGAGCTGAGAGAGCAGCTGAGAGAGCAGCGGAGCTGAGCGCAGCGAGAGAGGA 1532
1818 GCAGAGATGACAGCAGAGATGACAGCAGCAGAGATGACAGCAGAGATGACAGCA 1877
1533 GAGAGAGAGAGATGAGAGGCGGAGCGGAGAACTGGAACTGGAAAGGAGCGACA 1592
1878 GCAGAGATGACAGCAGAGATGAGAGCAGCAGAGATGACAGCAGAGATGAGAGCA 1937
1593 ACTTGATGGAGACGAGACCGGAGACAGAGACTCTGATTCAGAGAAAGAGAGCAGA 1652
1938 GCAGAGATGACAGCAGAGATGAGAGCAGCAGAGATGACAGCAGAGATGAGAGCA 1997
1653 GGGCACCCTGCTCTGAGGAGCAGAGAGAGCTTGAAGTGTGAGTGAAGCTCTGA 1712
1998 GCAGAGATGACAGCAGAGATGAGAGCAGCAGAGATGAGAGCAGAGATGAGAGCA 2057
1713 TGACAAAAGCATCAGCTAGAGAGAACTTCAGAGATTCAGAGTGTGACTGGCAACCA 1772
2058 GCAGAGATGACAGCAGAGATGAGAGCAGCAGAGATGAGAGCAGAGATGAGAGCA 2117
1773 GAGGCAAGAAATGAGAGCAGAGAAAGTTCAGAGCTTAAGATTCAGAAATTCACCA 1832
2118 GAGCAGCAGAGATGAGAGCAGAGAGAGATGAGAGCAGAGATGAGAGCAGAGCA 2177
1833 CTTCACAGCAGAGATGAGAGCAGAGAACTTCAGAGCTTGAAGACTTATTCAGAGAAAC 1892
2178 GATGAGAGCAGCAGAGATGAGAGCAGAGCAGAGATGAGAGCAGAGATGAGAGCA 2237
1893 GATATCTAGTACCACTTAAACAAGTCCAGCAGAAAGTTTGCATAGAGACTTCCTTCT 1952
2238 GCAGCAGCAGCAGAGATGAGAGCAGAGAGAGAGAGCAGAGCAGAGAGAGAGCA 2297
1953 TACCTCAAAAGAGCTTGAAGCAAAAGAGCTGCGCCGCGCAGCAGCTCCGGAGAGCT 2012
2298 GCAGAGATGAGAGCAGAGAGAGAGATGAGAGATGAGAGCAGAGATGAGAGCA 2357
2013 GAGCAGAGTGAAGAGAGCAGAGTCAAAAGTTCAGAGAGATGATGTTTCAACAACA 2072

2358 GGAGCAGAGATTAGAGAGCAGAGAGAGTTAGAGAGCAGAGCAGAGATTAGAGCA 2417
2073 GCTGAAGAACTGAGAGAGATCATAGCAAAACAGCACTCCAGAAAGCAGAGCTCCCTGA 2132
2418 GCAGAGCAGAGATTAGAGAGCAGAGAGAGATTAGAGAGCAGAGCAGAGATTAGCA 2477
2133 GCGAGCGGAGCTGAAGCAGAAAGAGCAGAGAGAGCTGAGATTAGAGCAAAA 2192
2478 GAGCAGAGCAGAGATTAGAGAGCAGAGAGAGATTAGAGAGCAGAGAGTGAAGA 2537
2193 GGAAGCGCTCAGAGAGAGATTCAAGAAAGGAGCAAGCAATGCTGAGCATGTGACAGA 2252
2538 GCAAGAGCAGAGAGTGAAGAGCAAGAGAGAGAGAGAGAGAGAGATTAGAGAGGT 2597
2253 GAGAGAGCAGCAGCGCCCGGAGAAACCCAGAGAGAGCAGAGAGAGAGAGAGAGAG 2312
2598 GAGAGAGAGAGCAGAGAGCAGAGAGAGAGAGAGAGAGAGATTAGAGAGTGAAGA 2657
2313 TGTCAAGAGAGAGAGCGGAGAGAGAGCAAGCGGAGAAATGCAAGCAAGCAGAGTGG 2372
2658 GCAGAGAGCAGAGATTAGAGAGAGTGAAGAGAGAGAGAGAGAGATTAGAGAGGT 2717
2373 GCTTTTCATCCGATCAGAGAGCAGCTAAGCTGGCCACCCAGG 2416
2718 GGAAGAGCAGAGCAGAGAGATTAGAGAGAGTGAAGAGAGG 2761

RESULT 13
US-09-298-568-1
Sequence 1, Application US/09298568
Patent No. 6322792
GENERAL INFORMATION:
APPLICANT: Kieff, Elliott D.
APPLICANT: Ballester, Mary E.
APPLICANT: Kaye, Kenneth M.
TITLE OF INVENTION: RHADINO VIRUS LANA ACTS IN TRANS ON A UNIT OF RHADINO
FILE REFERENCE: 16412-10001R
CURRENT APPLICATION NUMBER: US/09/298,568
EARLIER FILING DATE: 1999-04-21
EARLIER APPLICATION NUMBER: US 60/109,422
EARLIER FILING DATE: 1998-11-19
NUMBER OF SEQ ID NOS: 3
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 3489
TYPE: DNA
ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-09-298-568-1

Query Match 2.8%; Score 144; DB 4; Length 3489;
Best Local Similarity 45.1%; Pred. No. 4e-27;
Matches 534; Conservative 0; Mismatches 650; Indels 0; Gaps 0;

1233 GGATCAGAGGCTGCTGAGGAGCGCTGTCAGAGGATGACAGCAGCCAGAGAAACT 1292
1578 GCAGAGGCCACAGCAGAGGAGCCACAGCAGAGGCCACAGCAGAGGAGCCACAGCA 1637
1293 GCCTGTGACATTTGAAGATAAAGCGGAGAACTTTCAGCGAGGAGTGTGAGCTGGA 1352
1638 GCAGAGGCCACAGCAGCGGAGCCACAGCAGAGGCCACAGCAGAGGAGCCACAGCA 1697
1698 GCGGAGGCCACAGCAGCGGAGCCACAGCAGCGGAGCCACAGCAGCGGAGCCACAGCA 1757
1413 GAGCGCGCCGCGCAGAGAGAGAGAAAGCGGAGCGCAGAGAGAGAGCGGTTGCTCACT 1412
1758 GCGGAGGCCACAGCAGCGGAGCCACAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1472
1473 GCTGAGCTGAGAGAGCAGCTGAGAGAGCGGAGCTGAGAGCGGAGCGAGAGAGAGGA 1532

```

Db      1818 GCAGGATGAGCAGCAGCAGATGACAGAGGATGAGCAGCAGATGAGCAGCA 1877
Qy      1533 GAGGAGGAAGGATCGAGAGCGCGAGCCGCAAAACGGGAATCTGGAAAGCAGCCCA 1592
Db      1878 GCAGGATGAGCAGCAGCAGATGACAGCAGCAGATGAGCAGCAGATGAGCAGCA 1937
Qy      1593 ACTTGATGGGAAACGGAAACCGGAGCAGGAACTCTTGATCGAGGAAACGAGCAGCA 1652
Db      1938 GCAGGATGAGCAGCAGCAGATGAGCAGCAGCAGATGAGCAGCAGATGAGCAGCA 1997
Qy      1653 GGGGACCGGTGTCTTGAAAGCAGAGGAGAAGACTCTGAGTTTGAATGAAAGCTTGAA 1712
Db      1998 GCAGGATGAGCAGCAGCAGATGAGCAGCAGATGAGCAGCAGATGAGCAGCA 2057
Qy      1713 TGACAAAAGCCTCAGCTAGAGGAAACTTCAAGATATCAGGTCTCACTGGCAGCCA 1772
Db      2058 GCAGGATGAGCAGCAGCAGATGAGCAGCAGATGAGCAGCAGCAGCAGATGAGCA 2117
Qy      1773 GAGGCAAGAAATTGAGAGCAGCAAGCTCTAGAGCTTGAAGTTGCTGAAATCAGCCA 1832
Db      2118 GCAGCAGCAGATGAGCAGCAGCAGCAGATGAGCAGCAGCAGATGAGCAGCAGCA 2177
Qy      1833 CTTCAGCAGCAGCTTGAGAGAACTCTAGCAAAATGCTTGAGAGACTTATTCAGAGAAA 1892
Db      2178 GGATGAGCAGCAGCAGCAGATGAGCAGCAGCAGATGAGCAGCAGCAGCAGATGAG 2237
Qy      1893 GATATCTCAGTGCAGCTTAAACAAAGTCCAGCAGAACTTTGCTATAGACTCGCTTCT 1952
Db      2238 GCAGCAGCAGCAGATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 2297
Qy      1953 TACCTCAAAAAGCCTTGAGAGCAAAAGAGCTGGCCCGCAGCAGCTCCGGAGCAGCT 2012
Db      2298 GCAGGATGAGCAGCAGCAGCAGCAGATGAGCAGCAGCAGCAGATGAGCAGCA 2357
Qy      2013 GAGCAGGCTGAGAGAGAGCAGCAGCTCAAGCTGAGAGATGATGTTTTCACAAACA 2072
Db      2358 GAGCAGCAGGATGAGAGCAGCAGCAGCAGCAGATGAGCAGCAGCAGCAGATGAG 2417
Qy      2073 GCTGAAGAACTGAGAGAGATCATAGCAAAACAGCACTCCAGAGCAGAGTCCCTGGA 2132
Db      2418 GCAGGAGCAGGATGAGAGCAGCAGCAGCAGCAGATGAGCAGCAGCAGCAGATGAG 2477
Qy      2133 GGCAGCGCAGCTGAAAGCAGAAAGCAGAGAGAGGAAAGCTTGAGTTAGAGAACAAA 2192
Db      2478 GCAGCAGCAGCAGCAGATGAGAGCAGCAGCAGCAGATGAGCAGCAGCAGAGTGA 2537
Qy      2193 GGAAGACGCTCAGAGCAGATTCAGGAAAGGAGCAAGCAATGCTGAGCATGTGCAGCA 2252
Db      2538 GCAGAGCAGCAGAGTGAAGAGCAGAGCAGCAGCAGCAGCAGCAGCAGATGAG 2597
Qy      2253 GAGAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2312
Db      2598 GAGAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2657
Qy      2313 TGTCAAGAAAGAGAGGAGCAGAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2372
Db      2658 GCAGAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2717
Qy      2373 GCTTTTCATCCGATCAGAGCCAGCTTAAGCTGGCCAGCCAGG 2416
Db      2718 GGAAGAGCAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2761

```

```

RESULT 14
US-09-410-399-1
; Sequence 1, Application US/09410399
; Patent No. 6482587
; GENERAL INFORMATION:
; APPLICANT: Robertson, Erle S.
; TITLE OF INVENTION: Methods to Inhibit or Enhance the Binding of Viral DNA
; TITLE OF INVENTION: to Genomic Host DNA
; FILE REFERENCE: US-03778
S

```

```

; CURRENT APPLICATION NUMBER: US/09/410,399
; CURRENT FILING DATE: 1999-10-01
; NUMBER OF SEQ. ID NOS: 6
; SOFTWARE: Patent Ver. 2.0
; SEQ ID NO 1
; LENGTH: 3489
; TYPE: DNA
; ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-09-410-399-1

```

```

Query Match      2.88; Score 144; DB 4; Length 3489;
Best Local Similarity 45.18; Pred. No. 4e-27;
Matches 534; Conservative 0; Mismatches 650; Indels 0; Gaps 0;

Qy      1233 GCATCAGAGGTGCTTGAGAGCCGCTGTCAAGGATGAGCAGCAGCAGGAAAGT 1292
Db      1578 GCAGAGCCACAGAGCAGAGAGGAGCCACAGCAGCAGAGCCACAGCAGAGAGCCACAGCA 1637
Qy      1293 GCTGTGACATTTGAGATTAAGAGCCGGAGAACTTGAGCCGAGCAGTGTGAGCTGGA 1352
Db      1638 GCAGAGCCACAGAGCAGCAGGAGCCACAGCAGCAGCAGCAGCAGCAGCAGCA 1697
Qy      1353 GAAGCCGCCAGCAGCCTCTTGAGAGCAGCAGCCGAAAGAGCAGAGCGCTTGCTCAGCT 1412
Db      1698 GCGGAGCCACAGCAGCAGGAGCCACAGCAGCAGCAGCAGCAGCAGCAGCAGCA 1757
Qy      1413 GAGCCGCGCGAGAGAGAGGAAAGAGCCGGAGCCGAGAGCAGAGGCGCAAGCCGCA 1472
Db      1758 GCGGAGCCACAGAGCAGGAGCCACAGCAGCAGCAGATGAGCAGCAGCAGATGAGCAGCA 1817
Qy      1473 GCTGAGCTGAGAGAGCAGCTTGAGAGAGCAGCAGCAGCAGCAGCAGCAGCAGAGAGGA 1532
Db      1818 GCAGAGTGAAGCAGCAGCAGATGAGCAGCAGCAGCAGATGAGCAGCAGCAGATGAGCAGCA 1877
Qy      1533 GAGGAGGAAGAGATCGAGAGCCGAGCCGAGGAGGAGGAGTGAAGTGAAGAGCAGCA 1592
Db      1878 GCAGAGTGAAGCAGCAGCAGATGAGCAGCAGCAGATGAGCAGCAGCAGATGAGCAGCA 1937
Qy      1593 ACTTGAAATGGGAAGGAAACCGGAGCAGAGAACTCTGAATCAGAGGAAACAGAGCAGGA 1652
Db      1938 GCAGAGTGAAGCAGCAGCAGATGAGCAGCAGCAGATGAGCAGCAGCAGATGAGCAGCA 1997
Qy      1653 GAGCAGCGTGTCTTGAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1712
Db      1998 GCAGAGTGAAGCAGCAGCAGATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGTGA 2057
Qy      1713 TGACAAAAGCAGTACGCTAGAGGAAACTTCAGAGATATCAGGTGTGATCTGCAACCCA 1772
Db      2058 GCAGAGTGAAGCAGCAGCAGATGAGCAGCAGCAGCAGATGAGCAGCAGCAGCAGTGA 2117
Qy      1773 GAGCAGAAATTTGAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1832
Db      2118 GAGCAGCAGATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 2177
Qy      1833 CTTCAGCAGCAGCTTGAGAGAACTCTAGCAAAATGCTTGAGAGACTTATTCAGAGAAA 1892
Db      2178 GGAATGAGCAGCAGCAGCAGATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGTGA 2237
Qy      1893 GATATCTCAGTACAGATTAAGAAACAAAGTCCAGCAGAACTTTGCTATGAGACTCGCTTCT 1952
Db      2238 GCAGCAGCAGCAGATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 2297
Qy      1953 TACCTCAAAAAGCCTTGAGAGCAAAAGAGCTTGCCCGCAGCAGCTCCGGAGCAGCT 2012
Db      2298 GCAGAGTTTGAAGAGCAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 2357
Qy      2013 GAGCAGGCTGAGAGAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2072
Db      2358 GAGCAGAGATTTAGAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2417
Qy      2073 GCTGAAGAACTGAGAGAGATATATAGCAAAACAGCACTCAGAGAGCAGAGCTCCCTGGA 2132
Db      2418 GCAGAGCAGAGATTTAGAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2477

```


Fri Dec 5 09:46:38 2003

us-09-674-237a-1.rni

Page 12

```

Qy      2373  GCTTTTCCATCCGATCAGAGGCCAGCTAAGCTGGCCACCCAGG 2416
          |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      19279 GGAAGAGCAGAGGACAGCAGAGTTAGAGAGGTGGAAGAGCAGG 19236

```

Search completed: December 5, 2003, 00:46:34
Job time : 229,815 secs

OY	466	GATGAAGAAGATGGAATCAAGTGGAAATTTTCATAGACCAATGAAGCTTATCAAACTGAAGCTA	525
Db	249	GATGGGAAGATGGAACACAGAAAGTCTCTATAGCTATGAAACTATCAAGTTAAAGTTG	308
OY	526	CAAGGATATCAAGTCTCCCTCCACACTTCCCTCTGATGAAACAGCAACCAAGT--GCT	582
OY	583	ATTTCACATGACCAAGCAATTTGGTATPAGAGGATTTGGTAGCATG-----CCACA	633
Db	369	CCACTAATCTCTGCTCGTTTGGATGGGAGACATGCCAATCTGTCCATTTCATATAGCTA	428
OY	634	CTCAGAGCTGTGCTCTCTGTGCCA-----TGGGCTCAAT	669
Db	429	TTGCTCCAGTTGGACCTATAGCAACACCCTTGCTCTTGCTACTTACAGGGACCAAGTAT	488
OY	670	CCAAGTTTGGAAATGCTCCACCCCTTAGTATTTCTGTGCCCTCAGCAGCAAGTGGCTCC	729
Db	489	CCTCCCTTAATGATAGCTGCTCCCTAGGCTCTTGTTAGTACATCTCATTTAACAAAT	548
OY	730	CTGGCTTAAGGGGCTCTCCCGTCAATACAGCTCTGCGCTGCTTGGCGCATCTCAGCC	789
Db	549	GGAACTGGCAAGTCTATTCAGGCTTATTCATTTCTTATCTTCTTCAACATTTGGCTCAT	608
OY	790	ACATGGCCAAAGATTTCTTCTTTCAGCAGATCTGATCCAGGGTCAATTAACATACTAAG	849
Db	609	GCATCATCTTACAGGCTGATGATGGGAGATTTGGTGGTGATCATGATCCAGAAAGCCAG	668
OY	850	TTACGAAGGCGCAATTCATTCGATGCGCACGGCCCTCCAG-----	892
Db	669	TCTCTGATTTGATTTAGATCTAGTACTCAATCTTCTCAACTGCTTCCCTCTCAGGAAAC	728
OY	893	-----CAGCAGAAATGGGCTGTGCTCACTACATCAAGGCTGAATATACAG	936
Db	729	TCACCTAAGACAGGGACCTCAGAGTGGGGAGTTCTCCAGGCTTCAAGATTAAGTATCGG	788
OY	937	CAGTTATTCACAGCCACACAAAACTATGATGAGCACTTAAACAGTCCCAAGCCAGA	996
Db	789	CAAAAATTTAAATGCTATGACAAAAAGCAAGACGGATACCTTCAGGTTTTCMACTGAA	848
OY	997	ACTATTTCTCATCAATCAAGTTTACCCCAAGGCTCAGCTGGCTTCAATATGAATCTTCT	1056
Db	849	AATGCCCTTCTTCAGTCAAAATCTCTCTCAAACTCAGCTAGCTACTATTTTGGACTGGCT	908
OY	1057	GACATTGATCAAGATGAAAACTCACTGCGAAGAAATTTATCTTACTATGACCTAAT	1116
Db	909	GACATCTGATGCTGAGCGGACAGTTGAAGCTGAAGATTTATTTCTGGCGATGCACCTCACT	968
OY	1117	GATGTTGCGATGCTGTCGTCAGCACGTCGGCCGCTGCTCCGCAATATACATCCCTCCT	1176
Db	969	GACATGGCCAAAGCTGACAGCCACTTACACTGACGTTGGCTCCCGAGCTTGTCCCTCA	1022
OY	1177	TCTCTTCAAGAAAGTTCGCTCCGCGACAGTGGGATGTCCTGTCATAAGCTCTTCTCTGTGAT	1236
Db	1029	TCTTTCAAGG-----GGGAAACAGAGTTGATTTCTGTAT	1066
OY	1237	CAGAGGCTGCTCTGAGAGCCGCTGCTCAGAGATGACACGACCCAGAGAAATCTGCT	1296
Db	1065	GGAACTCTGGCTTTCATATTCAGAAAAACACAAGAG--AAGACCTCTAGAAAGAACTGGCA	1122
OY	1297	GTGACATTTGAAGATPAAGAAAGGGGGAACCTTCAAGCGAGCGAGTGTGGAGCTGGGAAG	1356
Db	1122	GTTACTTTTGGAGACAAACGGAAGCCACTATGAACGAGAAACATGGAGCTTGGAGAG	1188
OY	1357	CGCGCCCAAGCGCTCTTGGAGCAGCAACGCAAAAGACAGAGACGGTGGCTCAGCTGAG	1416
Db	1182	CGACGCCAAAGTGTGATGAGCAGCAGCAGAGAGGAGGCTGAACGCAAAAGCCAGAAAGAG	1244
OY	1417	CGCGCCGAGCAGGAAGAAAGAGGGGAGCCGCTCAGAGCAGAGAAGCCCAAGCGGCAAGCTG	1476
Db	1242	AAGGAAGGTGGAGCGGAAACAGAGAACTGCAAGACCAAGATGGAAGAAAGCAGCTG	1301

1477	GAGCTGGAGAAAGCAGCTCGAGAAACACACGGGAGCTGGAGCGGCAGCGAGAGAGAGAGG	1536
1302	GAGTTGGAGAAACCTCTGGAGAAACACAGAGAGAGCTGGAGAGACACGGGAGAGAGAGG	1361
1537	AGGAGAGAGATCGAGAGACCGCGAGCGCGCAAAACGGGAACTGGAAAGCGACGCAACT	1596
1362	AGAAAGGAGATAGAAAGACGAGAGAGCGAGCAAAAACAGGAGCTTGAGAGCAACGCCGTTTA	1421
1597	GAATGGGAAACGGAAACCGGAGACAGGAACTCTGAAATCAGAGAAACAAAGAGCAGAGAGGC	1656
1422	GAATGGGAAACCTCGCTCGGACAGAGAGCTGCTCAGTACAGAGAGACAGGAGAAACAAAGAGAC	1481
1657	ACCGTGGCTCGAAGGCAAGGAGGAAGACTCTGGAGTTTGAGTGAAGAGCTGTAATGAC	1716
1482	ATTGTGAGCTGAGACTCCAGAAAGAAAGTCTCCACTGGAACTGAAAGCAGTAAATGGA	1541
1717	AAAAAGCATCAGCTAGAAAGGAAAACTTCAGAGATATCAGAGTGTGCACTCGGACCCAGAGG	1776
1542	AAACATCAGACGAGATCTCAGGACGACTACAGATGTCCAAATTCABAAAACCAACCAAAAG	1601
1777	CAAGAAATTGAGACACGAAACAGTCTAGAGAGCTAAGAAATTGCTAAATCACCACCTTA	1836
1602	ACTGAGCTAGAACTTTTGGATTAACACAGTGTGACCTGGAAATTATGGAATCAACCACTT	1661
1837	CAGAGAGAGTTGGAGGAATCTCAGCAATGCTTGGAAAGACTATTTCCAGAGAAACAGATA	1896
1662	CAACAGAGCTTAAGAAATATCAAAATTAAGTTATCTATCTGGTCCCTGAGAAACAGCTA	1721
1897	CTCAGTGCACGTTTAAAAACAAGTCCAGCAGAACAGTTTGATAGAGATCGCTCTTAAC	1956
1722	TTTAAACGAAAGAAATTTAAAAACATGACAGCTCAGTAAACAACCTGATCAGGGATCAGTTTA	1781
1957	CTCAAAAGAGCTTTGGAAGCAAAAGAGCTGGCCCGGACAGCAGCTCCGGAGACAGCTGGAC	2016
1782	CTTCATAAAAAGTCATCAGAAAAAGAGAAATTTATGCCAAAGACTTAAGAACCAATTAGAT	1841
2017	GAGGTGAGAGAGAGACAGCTCAAGCTCAAGCTGAGAGATTAATGTTTCAACAACACAGCTG	2076
1842	GCTCTTGAAGAAAGAACTGCATCTCAAGCTCTCAAAAATGATTCATTAAACATCACACTG	1901
2077	AAGGAAGTGAAGAGATATCAGTACAAACAGCAATCCAGAAAGCAGAGAGCTCCCTGGAGCA	2136
1902	AAGGAATCTCAGAGAAAGCTATTAATACACAGCAGTTAGCCCTTGAAACAATTCAATTAATC	1961
2137	GCAGGACTGAAGCAGAAAGAGCAGAGAGAGAGAGAGCTGAGTTAGAGAAACAAA	2192
1962	AAACGTGACAAATTGAAGAAATGAAAGAAAGAAAGATTAGAGCAAAAAAAAAAAAAA	2017
RESULT 2		
US-09-907-969-72		
; Sequence 72, Application US/09907969		
; Publication No. US20030091580A1		
GENERAL INFORMATION:		
APPLICANT: Mitcham, Jennifer L.		
APPLICANT: King, Gordon E.		
APPLICANT: Algate, Paul A.		
APPLICANT: Fliny, Steven P.		
APPLICANT: Retter, Marc W.		
APPLICANT: Ranger, Gary Richard		
APPLICANT: Reed, Steven G.		
APPLICANT: Vedvick, Thomas S.		
APPLICANT: Carter, Darlick		
APPLICANT: Hill, Paul		
APPLICANT: Albone, Paul		
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY		
TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER		
FILE REFERENCE: 210121.462C8		
CURRENT APPLICATION NUMBER: US/09/907,969		
CURRENT FILING DATE: 2001-07-17		
NUMBER OF SEQ ID NOS: 596		
SOFTWARE: FastSeq For Windows Version 4.0		
; SEQ ID NO 72		

LENGTH: 2017
 TYPE: DNA
 ORGANISM: Homo sapiens
 US-09-907-969-72

Query Match 9.5%; Score 484.8; DB 11; Length 2017;
 Best Local Similarity 55.4%; Pred. No. 1.1e-128;
 Matches 1128; Conservative 0; Mismatches 812; Indels 96; Gaps 6;

226 GGGGTGATCAGAGGTGAACGTATAGAACATGGCTCAGTTTCCACACTTTCGGT 285
 9 GGTGAGAGCTGCAAGAGAGTACAGATCATATGCTCAGTTTCCACAGAGTAAT 68
 286 GGTAGCTGGATGTCTGGCCATACTGTGAGAGAGGCGCAAGCATGACAGATTC 345
 69 GAGAGGCCAAATATGTGGGTATTAATCTGAGAGAAAGTATGATGATTAACGATT 128
 346 CTTAGCCTGAGACCGATAGCGGATTTATTAATCTGTGATCAGCGAGAACTTTTTC 405
 129 GATTAACCTCAACCTTCAGGAGGTTACATACAGTGATCAAGCCGTAATTTTCCCTA 188
 406 CAATCTGGGTACCTCAGCTGTCTTACGACAAATATGGCGGTAGCGACATGATTAAC 465
 189 CAGTCAGGCTGTGCGGCGCCGATTAGTGAATATGGCCCTTATCAGATCTGAACAA 248
 466 GATGAGAGATGATCAAGTGAATTTTCCATAGCCATAGAGCTTATCAACTGAAGCTA 525
 249 GATGAGAGATGATGACAGCAAGAGTTCTTATAGCTATGAAATCTCATCAAGTTAAAGTTG 308
 526 CAAGATATCAGTCCCTCCACACTCCCTGTGATGAAACAGCAACCAAGT--GCT 582
 309 CAGGCGCAACAGTGTGCTGATGATCTCCCTCATATGAAACCAACCCCTATGTTCTCT 368
 583 ATTTCAGAGCACACAGATTTGGTATAGAGAGATTCGTAAGATG-----CAACA 633
 369 CCACTAATCTGTGCTGTTTGGATGAGAGATGCCAATCTGTCAATTCATGACCA 428
 634 CTCACAGCTGTGCTGCTGACCA-----TGGGCTCATTT 669
 429 TTGCCCTCAGTTGACCTATAGACACACCTTGTCTTGTCTACTTACAGGACCAATAT 488
 670 CCAAGTTTGGATATGTCACCCCTTATGATCTTGTGCTCCCTCAGACAGAGTCCCTCC 729
 489 CTTCCCTATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 548
 730 CTGGCTAAGGAGCTCTCCCGTCAATACAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 789
 549 GGAACCTGCAATCTCAATCAGCTTATCCATTCCTTATCTTCTTCAACATTTGCTCAT 608
 790 ACATGSCAAAGAGTCTTCTCTCAGACAGTCTGTCAGAGTCAAGATTTAAACATTAAG 849
 609 GCATCATCTTACAGCTGATGATGAGAGATTTGGTGTGCTAGTATCCAGAAAGGCCAG 668
 850 TTACAGAGAGCAATCATTCATGATGTCGCCAGGCCCTCTCAG-----892
 669 TCTCTGATTTAGATCTAGTACTCAACTTCTCAACTGCTTCCCTCAGGGAAC 728
 893 -----CAGCAGATGGGCTGTGCTCTCACTCATCAAGGCTGAATACAG 936
 729 TCACCTTAGACAGGAGCTCAGAGTGGGAGTCTCAGCTTCAAGATTTAAAGTATCG 788
 937 CAGTTATTCACAGCAGCAAACTATAGTGAACATTTAAACAGTCCCAAGGAGAA 996
 789 CAAAAATTTAATGTCTAGCAAGAGCATGAGCGGATACCTCAGAGTTTCAAGCTAGA 848
 997 ACTATTTCTATGCAATCAAGTTTACCCAGGCTCAGCTGCTTCAATATGATCTTCT 1056
 849 AATGCCCTTCTTCAATCATCTCTCAAACTCAGCTAGCTATTTGGATCTGCGCT 908
 1057 GACATTGATCAAGATGAAAACTCAGCTGAGAAAGATTTATCTTACTGATGACCTAAT 1116
 909 GACATGATGATGACGAGCAGTTGAAGAGTGAAGATTTATCTGCGATGACCTCAGCT 968

1117 GATGTTGCCATGTCTGATCAGCCAGCTCGCCGCTCTGCTCCAGAAATACATCCCTCT 1176
 969 GACATGCGCAAGAGCTGACAGCCACTTACCACTGAGCTTCCCTCCGAGTTTCCCTCA 1028
 1177 TCCCTCAGAAAGTTCGCTCCGCAAGTGGATGTCCTCATTAAGCTCTTCTGTGAT 1236
 1029 TCTTTTCAGAG-----GGGAAACCAAGTTATCTGTAT 1064
 1237 CAGAGCTCCCTGAGAGAGCCCTGTCAGAGATGAGCAGCAGCCAGAGAAATCTGCT 1296
 1065 GGAACCTGCTTCAATATCAGAAAACACAAAG--AAGACCTCAGAAAGAACTGCCA 1121
 1297 GTGACATTTGAAGTAAAGAGCGGAGAACTTCAAGCAGAGCAGAGTGTGAGCTGAGAG 1356
 1122 GTTATCTTTGAGAGCAAAACGAAACCACTATGAAAGAGAAATGAGCTGAGAGAG 1181
 1357 GCGCCGCAAGCGCTCTTGGAGCAGAGCGCAAAAGACAGAGCGGTTGCTCAGCTGAG 1416
 1182 CGAGCCCAAGTGTGATGAGCAGCAGCAGAGAGAGCTGAACGCAAGCCAGAAAGAG 1241
 1417 CGCGCCGAGCAGAGAGAGAAAGAGCGGAGCGCCAGAGCAGAGAGCCAGCGCAGCTG 1476
 1242 AAGAAAGATGAGAGCGGAAACAGAGAACTGCAAGAGCAAGATGGAAGAGCAGCTG 1301
 1477 GAGCTGAGAAACAGCTGAGAGAGAGCGGAGCTGAGCGGCGAGCAGAGAGAGAGAG 1536
 1302 GAGTTGAGAAACGCTTGGAGAAACAGAGAGCTGAGAGACAGCGGAGAGAGAGAG 1361
 1537 AGAAGAGATGAGAGAGCGCGAGCGCCAAACCGGAACTGGAAGAGCGAGCAACTT 1596
 1362 AGAAAGAGATGAGAAAGAGAGAGCAGAGCAAAACAGAGCTTGAAGACAGCGGTTTA 1421
 1597 GAATGGAACGGAACCGGAGACAGAACTCTGATCAGAGAAACAGAGCAGAGAGGCT 1656
 1422 GAATGGAAGAGCTCGCTGCGCAGAGCTGCTCACTCAGAAAGACAGGAGAACAGAG 1481
 1657 ACCGTGCTCTGAGAGAGAGAGAGAGCTGAGAGTTGAGTGAAGCTCTGATGAC 1716
 1482 ATTGTCAGGCTGAGCTCAGAAAGAAAGTCTCCACCTGGAACCTGGAAGATGATGGA 1541
 1717 AAAAAAGCATCAGCTAGAAAGAAACTTCAAGATATCAGGTGTCAGCTGCAACCAAGAG 1776
 1542 AAACATCAGCAGATCTCAGGCGAGACTACAAAGATGCCAAATCAGAAAGCAAAAG 1601
 1777 CAGAAATTTGAGAGCAAGAAAGTCTGAGAGCTTAAGATTTGTAATCCACCTTA 1836
 1602 ACTGAGCTAGAAAGTTTGGATTAACAGTGTGACTGGAATTTAGAAATCAAACACTT 1661
 1837 CAGCAGAGTTGACAGAAATCTCAGCAAAATGCTTGAAGACTTATCCAGAGAAACAGATA 1896
 1662 CAACAAAGCTTAAAGAAATATCAAAATAGCTTATCTGTGCTCCGAGAGAGAGCTTA 1721
 1897 CTCACTGACCACTTAAACAAAGTCCAGCAGAAACAGTTTGCATAGAGACTCGCTTCAAC 1956
 1722 TTAACGAAAGATTTAAACATGACAGCTCAGTACACACCTGATTCAGGATCAGTTTA 1781
 1957 CTCAAAAGAGCTTGAAGCAAGAGAGCTGCGCCGCGCAGCAGCTCCGGAGAGCTGAC 2016
 1782 CTTCAATTAAGATCTATCAGAAAGAGAAAGATTTATCCAAAGACTTAAAGAACTTAGAT 1841
 2017 GAGGTGAGAGAGAGAGCAGGCTCAAGCTGAGAGAGTGAATGTTTCAACACAGAGCT 2076
 1842 GCTCTTGAAGAAAGAACTGCACTTAAGCTCTCAGAAATGATTTCAATTAATCAGCTG 1901
 2077 AAGAACTGAGAGATACATAGCAAAACAGCAATCCAGAAAGCAGAGCTCCCTGAGAGCA 2136
 1902 AAGAACTCAGAGAAAGCTATTAATACAGAGAGTTAGCCCTTGAACAACTTATTAATTC 1961
 2137 GCGCAGCTGAGAGCAAGAAAGAGCAGAGAGAGAGCTGAGTTTGAAGAGCAAA 2192
 1962 AAACGTGACAAATTTGAAGAAATCGAAAGAAAGATTTAGGCAAAAAA 2017

```

RESULT 3
US-09-827-271-72
; Sequence 72, Application US/09827271
; Publication No. US20030165504A1
; GENERAL INFORMATION:
; APPLICANT: Retzer, Marc W.
; APPLICANT: Fanger, Gary R.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.462C6
; CURRENT APPLICATION NUMBER: US/09/827,271
; CURRENT FILING DATE: 2001-04-04
; NUMBER OF SEQ ID NOS: 461
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 72
; LENGTH: 2017
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-827-271-72

```

Db 1782 CTTTCATMAAAAGTCATCAGAAAAGAAATTAATGCAAGACTTAAGAACAAATTAGAT 1841
Qy 2017 GAGGTGAGAGAGAGACCAAGCTCAGAGATGATGTTTCAACACAGCTG 2076
Db 1842 GCTCTGAAAAAACTGATCTTAAGCTCTCAGAAATGATTCATTTAACAATCAGCTG 1901
Qy 2077 AAGGAATGAGAGATACATAGCAAAACAGCACTCCAGAGAGAGATCCCTGAGAGCA 2136
Db 1902 AAGGAATCAGAGAAAGCTATATACACAGCACTTGAAGCTTGAACAATTCATTAATC 1961
Qy 2137 GCGCGACTGAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2192
Db 1962 AAGCTGACAAATTAAGAGAAATGCAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 2017

RESULT 4
US-10-198-053-72
; Sequence 72, Application US/10198053
; Publication No. US20030124140A1
; GENERAL INFORMATION:
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Retter, Marc W.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Hill, Paul
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.462C9
; CURRENT APPLICATION NUMBER: US/10/198,053
; NUMBER OF SEQ ID NOS: 624
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 72
; LENGTH: 2017
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-198-053-72

Query Match 9.5%; Score 484.8; DB 14; Length 2017;
Best Local Similarity 55.4%; Pred. No. 1.1e-128;
Matches 118; Conservative 0; Mismatches 812; Indels 96; Gaps 6;

Qy 226 GGGGTGATCAGCAAGTGAACGTAAATAGAACCATGCTCAGTTTCCACACCTTTCGGT 285
Db 9 GGGTGAAGCTGCAAGAAAGTCAAGATCATATGCTCAGTTTCCACAGAGATGAT 68
Qy 286 GGTAGCTGATGCTGGGCGCATTAAGTGAAGAAAGGCGCAAGCATGACAGCAGTTT 345
Db 69 GGGGGCCAAATATGGGCTATTACATCTGAAGAAAGTAAAGCATGATAAAGAGTTT 128
Qy 346 CTTAGCTGAAGCCGATACCGGATTTATTAATGCTGATCAAGCGAGAACTTTTTC 405
Db 129 GATTAACCTCAACCTTCAAGAGGTATCAATAAGGATCAAGCCGTTCTTTTCCCA 188
Qy 406 CAATCTGGGTAACTCAGCTGCTTATAGCACAAAATATGGGCGCTAGAGCATGAATAC 465
Db 189 CAGTCAAGTCTGCGGCGCCGTTTATGCTGAATATGGGCTTATCAAGTCTGAACAG 248
Qy 466 GATGAGAGATGATCAAGTGAATTTTTCATAGCCATGAGCTTAACTGAAGCTA 525
Db 249 GATGGAGATGAGACAGCAAGAGTCTCTATAGCTATGAACCTCAAGTTAAAGTTG 308
Qy 526 CAAGATATCAAGTCCCTCCACACTCCCGCTGATAGAAACAGACAGCAGT---GCT 582
Db 309 CAGGGCAACAGTGTCTGTAGTCTCTCCCTATATAGAAACACCCCTATGTTCTCT 368
Qy 583 ATTTCAAGTCAACAGCATTTGATAGAGAGATGCTAGCATG---CCACCA 633
Db 369 CCACTATCTCTGCTGTTTGGGATGGAGACATCCCAATGTCTCATATACAGCA 428
Qy 634 CTCAAGCTGTTCTCTCTGCA---TGAGCTTCAT 669
Db 429 TTGCTCTCAGTGTGACCTATAGCAACCCCTGTCTTCTGCTACTTCAAGGACCAATAT 488

Qy 670 CCAATGTTGATGATGCTCCACCTTATGATCTTCTGCTCCCTCAGACAGAGCTCC 729
Db 489 CTTCCCTTATATGCTGCTCTCTCTAGGCTTCTGTTAGTACATCTCATTTACCAAT 548
Qy 730 CTGGCTTAAGGGGCTCTCTCCGTCATACAGCTCTGCTGCTGCTGCTGCTGAGCC 789
Db 549 GGAATGCAAGCTCATTCAGCTTATTCATTCCTTATTTCTTCAACATTTGCTCAT 608
Qy 790 ACATGAGCAAAAGATTTCTCTTCAAGAGATCTGCTCCAGGGTCAACATTAACACTAAG 849
Db 609 GCATCATCTTACAGCTGATGATGAGAGATTTGTTGCTGATGATATCCAGAAAGCCAG 668
Qy 850 TTACAGAAAGCAAAATCATGATGATGCGCAGGCGCCCTCCAG----- 892
Db 669 TCTCTGATGATTTAGATCTAGTACCTTCACTCTTCACTGCTTCCCTCAGGGAAC 728
Qy 893 -----CAGCAGATAGGCTGCTCCTCAGTCAATCAAGCTGAATACAG 936
Db 729 TCACCTAAGACAGGAGCCTCAGAGTGGGAGTTCTCAGGCTTCAAGATTAAGTATGG 788
Qy 937 CAGTTATTCACAGCAAGCAAAACTATAGTGAACCTTAACAGTCCCAAGCAGAGA 996
Db 789 CAAAATTTAATATGCTAGACAAAGCATGAGCGGATCTCTCAGGTTTCAAGCTAGA 848
Qy 997 ACTATTTCTGATCATCAAGTTTACCCAGGCTCAGGCTTCAATATGAAATCTTCT 1056
Db 849 AATGCCCTTCTCAGCAAAATCTCTCAAACTCAGCTAGCTATATTTGAGCTCTGCT 908
Qy 1057 GACATTTGATCAAGATGAGAAACTCACTGAGAAATTTATCTGATGACCTAAT 1116
Db 909 GACATGATGATGAGAGACAGTGAAGTGAAGATTTATCTGCGCATGACCTCACT 968
Qy 1117 GATGTTCCATCTGCTGACCACTGCGCCGCTCTGCTCCAGAAATATCCCTCT 1176
Db 969 GACATGCGCAAAAGCTGACAGCCACTACCACTGACGTTGCTCCGAGCTTCTCTCA 1028
Qy 1177 TCCCTCAGAAAGTTCGCTCCGCGAGTGGAGATGCTCCATATAGCTTCTCTGAT 1236
Db 1029 TCTTTCAGAG-----GGAAAGCAAGTTGATTTCTGTAT 1064
Qy 1237 CAGAGGCTGCTCAGAGAGCCGCTCTCAGAGATGACAGCCAGAGAGAACTGCT 1296
Db 1065 GGAATCTGCTCTCATATCAAGAAACAGAGAG---AAGGCTCAGAGAGAACTGCCA 1121
Qy 1297 GTGACTTTGAAGATTAAGAGCGGAGAACTTCAAGCAGAGGAGCTGAGAGAG 1356
Db 1122 GTTACTTTTGAAGCAAAAGGAAAGCACTATGAACAGAGAAACATGAGCTGAGAG 1181
Qy 1357 GCGCGCAAGCGCTCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1416
Db 1182 CGAGCCAAAGTGTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1241
Qy 1417 CCGCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1476
Db 1242 AAGGAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1301
Qy 1477 GAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1536
Db 1302 GAGTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1361
Qy 1537 AAGAGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1596
Db 1362 AAGAAAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1421
Qy 1597 GAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1656
Db 1422 GAAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1481
Qy 1657 ACCGTGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1716
Db 1482 ATTTGCAAGCTGAGCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1541
Qy 1717 AAAAAAGATCAGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1776

Db	1542	AAACCTCAGCAGATTC	CAAGCAGATCT	CAAGATGTC	CAATCAGAAACCAAAAG	1601
Qy	1777	CAAGAAATTGAGAGCA	CAACAGCTTAGAGAGCT	TAAGAAATTGCTGAATCA	CCCACTTA	1836
Db	1602	ACTGAGCTAGAAATTT	TGGATTTAAACAGTGTGAC	CTGAAATTTATGGAATCA	CAACACTT	1661
Qy	1837	CAGCAGCAGTTGCAGG	AATCTTCAGCAAAATGCTT	GGAAACCTTATTCAGAGAAACAGTA	1896	
Db	1662	CAACCAAGGCTTAAAG	AAATATCAATTAAGCTTATCTATCTGGTCC	CTGAGAAAGCAGTA	1721	
Qy	1897	CTCAGTGCACGTTTAA	AAACAGTCCAGAGAAACAGTTTGATGCAT	TAGAGATCCGCTTCTTAC	1956	
Db	1722	TTAAACGAAAGATTT	AAAAACATGCAAGCTCAGTACAT	CACACCTGATTCAGGATCAGTTTA	1781	
Qy	1957	CTCAAAAGAGCCTTGA	AGCAAAAGAGCTGGCCCGCAGACAGCT	CCGGAGCAGCTGAGC	2018	
Db	1782	CTTCATAAAAAGTCA	TCAAGAAAGAGAAATTTATGCCAAAGACTT	AAAGAACATTAAT	1841	
Qy	2017	GAGGTGAGAGAGAG	ACCAAGCTCAAAAGCTGCAGAGATTTGATGTTT	CAACCAACCAAGCTG	2076	
Db	1842	GCTCTTGAAGAAAG	AAGTCACTTAAGCTCTCAGAAATGATTTATTTA	CAATCAGCTG	1901	
Qy	2077	AAGCAACTGAGAGAT	ACTATGCAAAACAGCACTCCAGAAACAGAGCT	CCTGAGGCA	2133	
Db	1902	AAGCAACTCAGAG	AAAGCTATATATACACAGCAGTTAGCCCTTGA	CAAACTTATTAATC	1961	
Qy	2137	GCGGCACTGGAAG	CAGAAAGACAGAGAGAAAGCCTGGAGTTAGAGAGCAAA	2192		
Db	1962	AAACCTGACAAATTT	AAGGAAATTCGAAAGAAAGAAATTTAGAC	CAAAAAAAAAAAAAA	2017	

RESULT 5

```

US-09-764-881-55
; Sequence 55, Application US/09764881
; Publication No. US2003012546A9
;
; GENERAL INFORMATION:
;
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PIZ07
; CURRENT APPLICATION NUMBER: US/09/764,881
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 192
;
; SOFTWARE: PatentIn Ver. 2.0

```

```

? SEQ_ID NO 55
? LENGTH: 568
? TYPE: DNA
? ORGANISM: Homo sapiens
? FEATURE:
? NAME/KEY: SITE
? LOCATION: (481)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: SITE
? LOCATION: (536)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: SITE
? LOCATION: (556)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: SITE
? LOCATION: (562)
? OTHER INFORMATION: n equals a,t,g, or c
US-09-764-881-55

```

Query Match	8.0%;	Score 406;	DB 11;	Length 568;
Best Local Similarity	86.2%;	Pred. No. 2.9e-106;		
Matches 480;	Conservative	0;	Mismatches 73;	Indels 4;
				Gaps 3;

[illegible][illegible]

RESULT. 6

US-09-879-957-193
; Sequence 193, Application US/09879957
; Patent No. US20020034755A1

APPLICANT: SPARKS, Andrew B.
HOFEMAN, No. US20020034755A1h

KAY, Brian K.
FOWLIKES, Dana M.
MCCONNELL, Stephen J.

TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND USING SAME

NUMBER OF SEQUENCES: 227

CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas

CITY: New York
STATE: New York
COUNTRY: USA

ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/879, 957
FILING DATE: 13-Jun-2001

CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/630,915

FILING DATE: 03-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie

REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-174

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 193:

SEQUENCE CHARACTERISTICS:
LENGTH: 2873 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: DNA
SEQUENCE DESCRIPTION: SEQ ID NO: 193:
US-09-879-957-193

Query Match 6.6%; Score 333.2; DB 9; Length 2873;
Best Local Similarity 61.0%; Pred. No. 1.1e-84;

Matches 633; Conservative 0; Mismatches 363; Indels 42; Gaps 4;

2867 CTCTCTGACCGTACTAGTGTGCGCAGTTACGGCAGAGATCAGCCCTTACCCGACCA 2926
541 CTTTTCACAACTTAATTAATATCATCATGCGAGAAAATCAGCCTTCACTCGAAGTG 600
2927 CAGCCACTGCTCTCTCCCATCTCCGCTCTGGCCAGGTTAAAAGTGAAGGCTTAC 2986
601 TGTCCCTGG---ATCTGATCACTTATCATGAGCAGGGGACAAAGTGTAAGAACTTAA 657
2987 AAGCGCAAGCCCTGTATCCCTGGAGAGCCAAAAGCAACCACTTAATTTTAAACAAA 3046
658 AAGCAGAGCCCTTTTCTCTGACGTGCAAGAAAGATTAACACTGAACCTTCTCAAAAC 717
3047 GTGACGTATCAACCGTCTGGAACAGCAAGACATGTGTGTGTTGGAGAAGTTCAAGTTC 3106
718 ATGACATTTATCTGTCTTGGAGCAGCAAGAAATGTGTGTGGGAGGTGATGAGAG 777
3107 AGAAGGTTGTCTCCCAAGTCTTACGTAACTCAATTCAGGCGCCGTGAAGAAATCCA 3166
778 GAAGAGATGTGTTCCCAATCTTATGTCAAGATCATCTCTGGAGATGAAGTAAACGGG 837
3167 CAAGCATGATACTGGCCCTACTGAAAGTCCGTGCTGTAAGAAAGATGGCTTCCCGG 3226
838 AA-----GAACGAAAGCTTTGTATGACAGCTGAATTAAGAAACCTTACCTCGG 885
3227 CCGCCAGGCAAGCATTTCCCGAGAGAGATTTATGTCATGTCACATGAGAGTTCTG 3286
886 CAGCCTATTCACT-----TGAAGAAATATATGTGACCTTATTCATATTCAGAGTGG 939
3287 AGCAAGAGATTTAACCTTTACAGCAAGGAGATGTGATGTGTACCAAGAAAGATGTG 3346
940 AACCTGGAATTTGACTTTCAAGAGATGAAGAAATATTGGTGACCCAGAAAGATGGAG 999
3347 ACTGTGAGACGGGAAGGTGGGAGCAAGTCCGGAAGTCTTCCCTTATATATGTAGAGC 3406
1000 AGTGTGAGACGGGAAGGTGGGAGATGAAGTGAATTTTCCATTAACCTATGTCAAC 1059
3407 TTAAGATTTACAGAGGCTCTGGAACGTGGGAAAACAGGAGATTTAGAGAAAAAATCTG 3466
1060 CAAGAGATCAAGAGATTTTGGAGGTGCTAGCAAGTCTGAGATCAATTAATTAACCTG 1119
3467 AAATTGCCAGGTTATGTCTTCTACGCTGCTATCTGTCTCCGAACAACTCACCTGGCTC 3526
1120 AATATGCTCAGGTAATCTTACGATATGTGTCTTCTGTCTGAACAACTTAAAGCTTGGAC 1179
3527 CTGGGAGCTGATTCGATTCGGAAGAAAGAACCCAGATGATGTGTGGAGAGAGATCTGC 3586
1180 CAGGAGATTAATTAATTTCTTAAGAAAAATTAAGAGTGTGTGGCAGAGAGATTTAC 1239
3587 AAGCTCGAGGAGAAAAAGCCAGATAGGTGTGTTTCAGCAAAATTAATGCAACTTCTAA 3646
1240 AAGCCAGAGAGAAAAAGCAGAGAAAGATGTGTCTCTGCACTCATGTTAACTTTGG 1299
3647 GCCCGGAAAACAGCAAAATTCACCCCAATCTGAGTCAACCAAGCCGAGTGCAGCCAGCAG 3706
1300 GTTCAAGTATGTGAAGAGCACAAGCTGCTTTCATC-----CTG 1338

3707 TGTGCAAGTATGCGGATGTACGATTACACGCCCGAAGCATGACGAAGTACGCTTCA 3766
1339 TATGTCAAGTATGCTATGTATGACTATGACGCAAAATATATAGATGAGCTTCACTTCT 1398
3767 GCMAAGCCAGATCATCAACGTCCTCAACAGAGAGACCCGAGCTGTGAAAGAGAG 3826
1399 CCAAGGCAACTCATTAATTAATGTAAGAACAAAGATGATCTGATGTGTGCAAGAGAGA 1458
3827 TCAGTGGGCAAGTGGGCTTCTCCATTCATTAATTAAGTGAAGTACACAGACATGACC 3886
1459 TCAAGGGGTGACTGTCTCTTCTTCAACTACGTTAAGATGACGACAGACTGATGC 1518
3887 CCAGCAGCAATGAATCA 3904
1519 CAAGTCAACATGACCCA 1536

RESULT 7

US-09-764-868-125
Sequence 125 Application US/09764868

Patent No. US2002016871A1

GENERAL INFORMATION:

APPLICANT: Rosen et al.

TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

FILE REFERENCE: PT232

CURRENT FILING DATE: 2001-01-17

NUMBER OF SEQ ID NOS: 1510

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 125

LENGTH: 4210

TYPE: DNA

ORGANISM: Homo sapiens

US-09-764-868-125

Query Match 6.5%; Score 331.4; DB 10; Length 4210;
Best Local Similarity 61.0%; Pred. No. 4.9e-84;

Matches 630; Conservative 0; Mismatches 361; Indels 42; Gaps 4;

2867 CTCTCTGACCGTACTAGTGTGCGCAGTTACGGCAGAGATCAGCCCTTACCCGACCA 2926
918 CTTTTCACAACTTAATTAATATCATCATGCGAGAAAATCAGCCTTCACTCGAAGTG 977
2927 CAGCCACTGCTCTCTCCCATCTCCGCTCTGGCCAGGTTAAAAGTGAAGGCTTAC 2986
978 TGTCCCTGG---ATCTGATCACTTATCATGACAGGAGCAAGTGTGAAGAACTTAA 1034
2987 AAGCCAGGCCCTGTATCCCTGGAGAGCCAAAAGACAAACCACTTAATTTTAAACAAA 3046
1035 AAGCAGAGCCCTTTGTCTCTGAGCTGCAAAAGAAATTAACCACTTGAACCTTCAAAAC 1094
3047 GTGACGTATCAACCGTCTGGAACAGCAAGACATGTGTGTGTTGGAGAAGTTCAAGTTC 3106
1095 ATGACATTTATCTGTCTTGGAGCAGCAAGAAATTTGGTGTGGGAGGTGCAATGGAG 1154
3107 AGAAGGTTGTCTCCCAAGTCTTACGTGAATCTGTAATTCAGGSCCCGTGAAGAAATCCA 3166
1155 GAAGAGATGTGTTCCCAATCTTATGTCAAGATCATCTCTGGAGTGAAGTAAACGGG 1214
3167 CAAGCATGATACTGGCCCTACTGAAAGTCCGTGATGTTAAAGAGATGGCTTCCCGG 3226
1215 AA-----GAACGAAAGCTTTGTATGACGCTGTAATTAAGAAACCTTACCTCGG 1262
3227 CCGCCAGGCAAGCATTTCCCGAGAGAGATTTATGTCATGTCACATGACAGATTTCTG 3286
1263 CAGCCTATTCACT-----GGAAGAAATATATGTGCACTTATTCATATTCAGATGTGG 1316
3287 AGCAAGAGATTTAACTTTACAGCAAGGAGATGTGATGTGTGTTACCAAGAAAGATGTG 3346
1317 AACCTGAGATTTGACTTTACAGAAAGTGAAGAAATATTGGTGTGACCCAGAAAGATGGAG 1376

QY 3347 ACTGTCGACGAGAAACGCTGGGCGACCAAGTCCGAGTCTTCCTCTTAATATGTGAGGC 3406
 DB 1377 AGTGTGACAGAGAGATTTGAGATAGAGTGAATTTTTCATCAAACTATGTCAAC 1436
 QY 3407 TTAAGATTCAGAGGCTCTGGAACCTGCTGGGAAAAACAGGAGTTTAGAAAAAAACCTG 3466
 DB 1437 CAAGAGATCAAGAGAGTTTGGAGTGTGAGCAATCTGAGCAATCAAAATAAAAAACCCTG 1496
 QY 3467 AAATTTGCCAGATTTATGCTCTCTACGCTGCTACCTGCTCCGGAACAATCAACCTGCTC 3526
 DB 1497 AGATTGCTCAGATTAATCTGACATATGTTGCTTCTGTTCTGAAACACTTGAACCTTTGAC 1556
 QY 3527 CTGGCAGCTGATCTGATCCGAAAAAAGAACCCAGTGTGATGTTGGAGAGAGAACTGC 3586
 DB 1557 CAGACACTTAATATTAATTTCTAAAGAAAAATACAGTGTGCTGCTGCGAAGAGAGTTAC 1616
 QY 3587 AAGCTGAGGAGAAAAAGCCGACATAGGCTGTTTCCAGCAATTAATGTCAACTTTCTAA 3646
 DB 1617 AGGCGAGAGAAAAAAGCGACAGAAAGAGTGTCTGCGACATGATTTAACTTTTGG 1676
 QY 3647 GCCCGGAAACAAGCAAAATCAACCCCACTGAGCTACCCAGACCGCACTGACGACGAG 3706
 DB 1677 GTCCAAAGTAGAAAGACCCACACTGCTTTTCATC-----CTG 1715
 QY 3707 TGTGCGAGGTATCGGATGTACGATTAACCGCCAGAACGATGACGAACCTAGCCTTCA 3766
 DB 1716 TATGTCAGGTATGCTATGATGATGATGACGAAATATGAGATGAGTCACTTCT 1775
 QY 3767 GCAAAAGCCAGATCATCAACGCTCTCAACAGAGAGACCCGAGCTGTGAAAGAGAGAG 3826
 DB 1776 CCAAGGACCAACTATGATGATGAACAAAGATGATCTGATGTTGGTGGCAAGAGAGA 1835
 QY 3827 TCAGTGGGCAAGTGGGCTCTTCCATCAATTAATGTAAGCTGACCAACATGAGACC 3886
 DB 1836 TCACAGGGGTGACGTGCTCTTTCTTCAAACTAGTTAAGATGACAGACTCAGATC 1895
 QY 3887 CCAGCCAGCAATG 3899
 DB 1896 CAAGTCACAGTG 1908

RESULT 8
 US-09-918-995-31258
 ; Sequence 31258, Application US/09918995
 ; Publication No. US20030073623A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Hyseq, Inc.
 ; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
 ; FILE REFERENCE: 20411-756
 ; CURRENT APPLICATION NUMBER: US/09/918,995
 ; PRIOR FILING DATE: 2001-07-30
 ; PRIOR APPLICATION NUMBER: US/09/235,076
 ; NUMBER OF SEQ ID NOS: 38054
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 31258
 ; LENGTH: 503
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; LOCATION: (1)...(503)
 ; OTHER INFORMATION: n = A,T,C or G
 US-09-918-995-31258

Query Match 6.3%; Score 322.6; DB 11; Length 503;
 Best Local Similarity 81.6%; Pred No. 3.7e-82;
 Matches 386; Conservative 0; Mismatches 84; Indels 3; Gaps 1;
 QY 1081 ACTGCAAGAAATTTATCTAGTATGACCACTTAATGATGTTGCCATGCTGTCAGCA 1140
 DB 30 AATTCAGAGAAATTTATCTGCAATGACCTCATGATGATGATGATGCTGCGCAACCA 89

QY 1141 CTGCGCCGCGCTGCTGCTCCAGAAATAGATCCCTCTCTTACAGAGAGTTGCTCCGCG 1200
 DB 90 CTGCGACCTGCTGCTCTCCAGATACATTCACCTTTTAAAGAGTTCTGCTGCGC 149
 QY 1201 AGTGGATGTCCTCATAGTCTTCTTCTGTGTGATCAGAGGCTGCTGAGAGCCGTG 1260
 DB 150 AGTGTATATCTGTCATAGTCAATCTGTATGATCAGAGGCTACCAAGAGAACAGTT 209
 QY 1261 TCAGAGATGACGACGAC---CAGAGAAAACTGCTGTGACATTTGAAGATTAAGAG 1317
 DB 210 TTAAAGATGAAACAACAATTAAGAAAGAAATTAACCTGTAAAGTTGAAGATTAAG 269
 QY 1318 CGGAGAACTTCAGCGAGGAGAGTGTGAGCTGAGAGAGCGCCAGAGCGCTTGTGAG 1377
 DB 270 CGGAGAACTTTGAACCTGCGACACTGGAAGAAAGAGAAAGCAAGCTCTTGGAA 329
 QY 1378 CAGCAGCGCAAGAGCAGAGCGGTTGCTCACTGAGCGCGCCAGCAGAGAGAGAAA 1437
 DB 330 CAGCAGCGCAAGAGCAGAGCGGCTGCGCCAGCTGAGCGCGGCGAGCAGAGAGAG 389
 QY 1438 GAGCGGAGCGCCAGAGCAGAGAGCCAGCGGACCTGAGAGCTGAGAGAGCAGTGGAG 1497
 DB 390 GAGCGTACGCGCCAGAGCAGAGCGCATTAAGCAACTGGAACCTGAGATGCAACTGG 449
 QY 1498 AAGCAGCGGAGCTGAGCGCGCAGAGAGAGAGAGAGAGAGAGATGCA 1550
 DB 450 AAGCAGCGGAGCTGAACCGCAGAGAGAGAGAGAGAGAGAGATTTGA 502

RESULT 9
 US-09-879-957-39
 ; Sequence 39, Application US/09879957
 ; Patent No. US2002003475A1
 ; GENERAL INFORMATION:
 ; APPLICANT: SPARKS, Andrew B.
 ; HOFFMAN, No. US2002003475A1h
 ; KAY, Brian K.
 ; FOWLES, Dana M.
 ; MCCONNELL, Stephen J.
 ; TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
 ; DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
 ; USING SAME
 ; NUMBER OF SEQUENCES: 227
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Pennie & Edmonds LLP
 ; STREET: 1155 Avenue of the Americas
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: USA
 ; ZIP: 10036-2711
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/879,957
 ; FILING DATE: 13-Jun-2001
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/630,915
 ; FILING DATE: 03-APR-1996
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Mastrock, S. Leslie
 ; REGISTRATION NUMBER: 18,872
 ; REFERENCE/DOCKET NUMBER: 1101-174
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212) 790-9090
 ; TELEFAX: (212) 869-8864/9741
 ; INFORMATION FOR SEQ ID NO: 39:
 ; SEQUENCE CHARACTERISTICS:

RESULT 10
 US-09-864-761-17127
 : Sequence 17127, Application US/09864761
 : Patent No. US20020048763A1
 : GENERAL INFORMATION:
 :
 : APPLICANT: Penn, Sharon G.
 : APPLICANT: Rank, David R.
 : APPLICANT: Hanzel, David K.
 : APPLICANT: Chen, Menseheng
 : TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
 : FILE REFERENCE: Aeomica-X-1
 : CURRENT APPLICATION NUMBER: US/09/864,761
 : CURRENT FILING DATE: 2001-05-23
 : PRIOR APPLICATION NUMBER: US 60/180,312
 : PRIOR FILING DATE: 2000-02-04

QY	2557	GTGATGTAAGGCCGAGCTGGAGAGCCAGAGATGGCTTGAGAGAGAGCTAAAGGGGAAAGCG	261
Db	1	GTGGATCAAAAGCCAAACTGGAGAACCCGGCTGGCTTGAGAGGAGAAATTAAGGAAAGAACGA	60
QY	2617	GGAGTGTTCCTCGAAACTATATGCAAGAAAATTCACAAAATATGATGTTCCCACTCCACCC	2674
Db	61	GGGTGTTCCTCGAAACTATGCAAGAAAATTCACAAAATATGATGTTCCCGCTCAAGTG	120
QY	2677	AAACCACTGACCGATCTGACATCTGCCCCCTGCCCAAATGCGCTCTTGCGTAGACCCCT	2734
Db	121	AAACCACTGACCTGATTTCAACATCTGCCCCCTGCCCAAATGCGCGCTTGCGTAGACCCCT	180

QY	2737	GCTCCTTTGCCAGATACCTCTTTCTGAGCCCTCAAAACCCCAACAACTGGGGAGACTTC	2756
Db	181	GCCCTTTGGGAGTAACCTCTTCAGAGCCCTCAAGACCCCAATAATATGAGGGCGACTTC	240
QY	2797	AGTTTCACAGT	2806
Db	241	AGCTTCACGT	250

```

RESULT 11
; US-09-864-761-26948
; Sequence 26948, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecmics-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 26948
; LENGTH: 286
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AP000311.1
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 0.87
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 32
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.66
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.69

```

```

1 OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1
2 OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.92
3 OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.98
4 OTHER INFORMATION: NT HIT: A2114488.1, EVALUE 0.006+00
5 OTHER INFORMATION: SWISSPROT HIT: Q15811, EVALUE 2.006-45
6 OTHER INFORMATION: EST_HUMAN HIT: AA773263.1, EVALUE 1.006-112
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013

```

Query Match	3.9%	Score 198.8	DB 9	Length 286
Best Local Similarity	87.2%	Pred. No. 1.7e-46		
Matches 218; Conservative	0	Mismatches 32	Indels 0	Gaps 0

Qy	2557	GTGATGAAAAGCCAGACTGGAGAGCCAGATGCTTGGAGAGAGCTGAAAGGGAAGACG	2616
Db	1	GTGATGAAAAGCCAAACTGAGGAACCCGGCTGCTTGGAGGAGATTTAAAGAAAGAACGA	60
Qy	2617	GGATGTTTCCCTGCAAACTATGCGAGAAAAGATTCCAGAAAATGAGTTTCCCACTTCAGCC	2676
Db	61	GGGAGGTTTCCCTGCAAACTATGCGAGAAAATCCCAAAAATGAGTTTCCCGCTCCAGGTG	120
Qy	2677	AAACCAAGTACCCGATCTGCATCTGCCCTGCCCCCAAACTGAGCTGCTGAGAACCCCT	2736
Db	121	AAACCAAGTACGATGTTCAACATCTGCCCCCTGCCCAAACTGGCTTGGCGTGAACCCCC	180
Qy	2737	GCTCCTTTGCGAGTGACCTCTTCTGAGGCCCTCCAAACCCCAACAACTGGGCGACCTTC	2796
Db	181	GCCCCCTTTGGCGAGTAACTTTCAGAGCCCTCCAGAACCCCTAATAAATCGGGCCGACTTC	240
Qy	2797	AGTCCACAGT	2806
Db	241	AGCTCCAGT	250

```

1  RESULT 12
2  US-09-864-761-30453
3  Sequence 30453, Application US/09864761
4  Patent No. US20020048763A1
5  GENERAL INFORMATION:
6  APPLICANT: Penn, Sharon G.
7  APPLICANT: Rank, David R.
8  APPLICANT: Hanzel, David K.
9  APPLICANT: Chen, Wensheng
10 TITLE OF INVENTION: HUMAN GENOME- DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
11 FILE REFERENCE: Aemica-X-1
12 FILE REFERENCE: Aemica-X-1
13 CURRENT APPLICATION NUMBER: US/09/864,761
14 CURRENT FILING DATE: 2001-05-23
15 PRIOR APPLICATION NUMBER: US 60/180,312
16 PRIOR FILING DATE: 2000-02-04
17 PRIOR APPLICATION NUMBER: US 60/207,456
18 PRIOR FILING DATE: 2000-05-26
19 PRIOR APPLICATION NUMBER: US 09/632,366
20 PRIOR FILING DATE: 2000-08-03
21 PRIOR APPLICATION NUMBER: GB 24263.6
22 PRIOR FILING DATE: 2000-10-04
23 PRIOR APPLICATION NUMBER: US 60/236,359
24 PRIOR FILING DATE: 2000-09-27
25 PRIOR APPLICATION NUMBER: PCT/US01/00666
26 PRIOR FILING DATE: 2001-01-30
27 PRIOR APPLICATION NUMBER: PCT/US01/00667
28 PRIOR FILING DATE: 2001-01-30
29 PRIOR APPLICATION NUMBER: PCT/US01/00664
30 PRIOR FILING DATE: 2001-01-30
31 PRIOR APPLICATION NUMBER: PCT/US01/00669
32 PRIOR FILING DATE: 2001-01-30
33 PRIOR APPLICATION NUMBER: PCT/US01/00665
34 PRIOR FILING DATE: 2001-01-30
35 PRIOR APPLICATION NUMBER: PCT/US01/00668
36 PRIOR FILING DATE: 2001-01-30
37 PRIOR APPLICATION NUMBER: PCT/US01/00663
38 PRIOR FILING DATE: 2001-01-30
39 PRIOR APPLICATION NUMBER: PCT/US01/00662
40 PRIOR FILING DATE: 2001-01-30

```


Db 181 AACTGAGAACCAATGTGAAAAAGCGCGGAGCTGAAACGCGAGAGAGAGAGAGAGA 240
Gy 1538 GGAAGAGATGAGAGCGCGCGAGCGCGCAAAACGCGA 1574
Db 241 GGAAGAAATTGAGAGCGCGAGGCTTACGACGCGCGAGA 277

RESULT 14
US-09-864-761-17644
Sequence 17644, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
FILE REFERENCE: Aecomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 17644
LENGTH: 263
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AP000049.1
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.3
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.2
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.3
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 3
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.9
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.2

OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.5
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 5.3
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.9
OTHER INFORMATION: EST_HUMAN HIT: BE542917.1, EVALUATE 3.00e-62
OTHER INFORMATION: SWISSPROT HIT: O35601, EVALUATE 2.80e-01
OTHER INFORMATION: NT HIT: AF114487.1, EVALUATE 1.00e-122
US-09-864-761-17644

Query Match 3.9%; Score 197.4; DB 9; Length 263;
Best Local Similarity 84.4%; Pred. No. 4e-46;
Matches 222; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

Gy 1298 TGACATTGAAGATTAAGACGGGAGAACTTCGACGAGCAGTGTGAGCTGGAGAGC 1357
Db 1 TAACTTTGAAGATTAAGACGGGAGAACTTTGACCTGGCAACTGGAAGAAC 60
Gy 1358 GCCGCCAAGCGCTCTTGAAGACGAGCGCAAGAGCAGAGCGGTTGCTCAGCTGAGC 1417
Db 61 GAAAGCAAGCTCTCTCTGAAACAGCAGCGCAAGAGCAGAGCGCTGCGCCAGCTGAGC 120
Gy 1418 GCGCCGAGCGAGAGAGAGAGAGAGCGGCGGAGCGAGAGCGGAGCGGAGCGAGCTGG 1477
Db 121 GGGCGAGAGAGAGAGAGAGAGAGCGGTAGCGCCAGAGCAAGAGCGCAAAAGCAACTGG 180
Gy 1478 AGCTGAGAGAGCAGCTGAGAGAGCAGCGAGCTGAGCGGAGCGAGAGAGAGAGAGA 1537
Db 181 AACTGAGAGCACTGGAAGAAAGCAGCGGAGCTAGAGACGAGAGAGAGAGAGAGA 240
Gy 1538 GGAAGAGATGAGAGCGCGGAG 1560
Db 241 GGAAGAAATTGAGAGCGCGAGAG 263

RESULT 15
US-09-864-761-10314
Sequence 10314, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
FILE REFERENCE: Aecomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30

```

1  PRIOR APPLICATION NUMBER: PCT/US01/00661
2  PRIOR FILING DATE: 2001-01-30
3  PRIOR APPLICATION NUMBER: PCT/US01/00670
4  PRIOR FILING DATE: 2001-01-30
5  PRIOR APPLICATION NUMBER: US 60/234,687
6  PRIOR FILING DATE: 2000-09-21
7  PRIOR APPLICATION NUMBER: US 09/608,408
8  PRIOR FILING DATE: 2000-06-30
9  PRIOR APPLICATION NUMBER: US 09/774,203
10 PRIOR FILING DATE: 2001-01-29
11 NUMBER OF SEQ ID NOS: 49117
12 SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
13 SEQ ID NO 10314
14
15 LENGTH: 480
16
17 TYPE: DNA
18
19 ORGANISM: Homo sapiens
20
21 FEATURE:
22
23 OTHER INFORMATION: MAP TO AP000311.1
24
25 OTHER INFORMATION: EXPRESSED IN B7474, SIGNAL = 0.87
26
27 OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.1
28
29 OTHER INFORMATION: EXPRESSED IN HEL100, SIGNAL = 1.3
30
31 OTHER INFORMATION: EXPRESSED IN FETAL LAYER, SIGNAL = 32
32
33 OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.66
34
35 OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.69
36
37 OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1
38
39 OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.92
40
41 OTHER INFORMATION: EXPRESSED IN ADULT LAYER, SIGNAL = 0.98
42
43 US-09-864-761-10314

```

Query Match	3.8%	Score 192	DB 9;	Length 480;
Best Local Similarity	81.6%	Pred. No. 2.2e-44;		
Matches 222; Conservative	0;	Mismatches 50;	Indels 0;	Gaps 0;

OY	2553	TGAGATCAACATCCGCGCAGAGATATATGTATGTGTGATGAAACCGACATCGAGAGACC	2582
Db	209	TTAGTTCCTACTTAAGCTGTGACTTTTCCACAGGTGATGAAACCAATCTGAGAAC	268
OY	2583	AGATGCGTTTGAGAGAGAGCTGAAAGGAGACGGAGTGTTCCTGCAAACTATGACGA	2642
Db	269	CGGCTGGCTTGAGAGAGAAATTTAAAAAGAAAGACGGGTGTTCCTGCAAACTATGACGA	328
OY	2643	AAAGATTCCAGAAAAATGAGTTCGCCACTCCACGCCAAACAGGACCGGATCTGACATCTGCG	2702
Db	329	GAAATTCCTCAAGAAATGAGGTTCCCGCTCCAGTGAACCAAGTGAATTCACACATCTGCG	388
OY	2703	CCCTGCCCCCAACTGAGCTTGTGCGTGAGAACCCCTGTCTCTTGTCCAGTGAACCTCTTCTGA	2762
Db	389	CCCTGCCCCCAACTGAGCTTGTGCGTGAGAACCCCGCCCTTGTGACATTAACCTTTTCAGA	448
OY	2763	GGCCTCCCAACCCCCCAACCAATGCGGACGACT	2794
Db	449	GGCCTCCCAACCCCTTAATTAATCTGGGCGGACT	480

Search completed: December 5, 2003, 06:15:39
Job time : 1089.45 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 4, 2003, 15:20:59 ; Search time 588.079 Seconds
(without alignments)
16717.740 Million cell updates/sec

Title: US-09-674-237A-2

Perfect score: 3642

Sequence: 1 atgctcagttccaccaccc.....tgaccaccagccagcaatga 3642

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_190903.*

```
1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*
25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3642	100.0	3723	21	AAZ39009
2	3639.4	99.9	5144	21	AAZ39025
3	3639.4	99.9	5738	21	AAZ39024
4	3630	99.7	5082	21	AAZ39008
5	2856	78.4	5195	20	AAZ34572
6	2831	77.7	5458	20	AAZ34571
7	2491.6	68.4	7435	23	AAZ34763
8	2429.6	66.7	5199	20	AAZ34570

Result No.	Score	Query Match	Length	DB ID	Description
9	2043.8	56.1	3319	23	ABK43498
10	2043.8	56.1	3466	22	AAI63825
11	1495	41.0	2079	20	AAZ34573
12	1483.8	40.7	2131	22	AAK94139
13	1291.4	35.5	2874	23	AAZ34572
14	1259.2	34.6	2131	22	AAH16578
15	1252.8	34.4	3231	20	AAZ34574
16	1104.2	30.3	1676	22	AAK94611
17	858	23.6	1389	17	AAZ39795
18	721	19.8	5828	24	AAH47247
19	713.8	19.6	6103	22	AAK52332
20	677	18.6	4625	21	AAZ39010
21	677	18.6	4975	21	AAZ39027
22	677	18.6	6014	21	AAZ39026
23	666.6	18.3	3593	21	AAZ39011
24	648	17.8	4447	22	AAZ39055
25	513.8	14.1	677	25	AAZ39055
26	485.4	13.3	877	22	AAK93179
27	485.4	13.3	877	22	AAK93179
28	477.4	13.1	2017	21	AAZ39011
29	477.4	13.1	2017	21	AAZ39011
30	387.4	10.6	1329	22	AAH15260
31	356	9.8	548	22	AAI80000
32	356	9.8	568	22	AAI63919
33	356	9.8	568	22	AAZ31621
34	356	9.8	568	22	AAK43814
35	340.8	9.4	2067	23	ABO55007
36	332.4	9.1	2873	17	AAZ39799
37	331.4	9.1	3746	23	AAK43586
38	331.4	9.1	4210	22	AAZ27090
39	281	7.7	831	22	AAH03435
40	260	7.1	676	22	AAH08146
41	258	7.1	747	17	AAZ39796
42	250	6.9	346	21	AAZ44038
43	210.8	5.8	395	23	AAZ84759
44	206	5.7	701	25	AAZ57579
45	198.8	5.5	270	22	AAZ46803

ALIGNMENTS

RESULT 1
AAZ39009
AAZ39009 standard; cDNA; 3723 BP.
XX
XX
AC AAZ39009;
XX
XX
DT 28-FEB-2000 (first entry)
XX
XX
DE Mouse Eesl coding sequence.
XX
XX
XX Mouse; murine; Eesl; Eesl; endocytosis; vesicular trafficking;
XX regulation; actin cytoskeleton; detection; cancer; infection;
XX EH-domain and SH3-domain regulator of endocytosis; anticancer;
XX antiproliferative; antiviral; ss.
XX
XX
OS Mus sp.
XX
XX
PN W0955728-A2.
XX
XX
PD 04-NOV-1999.
XX
XX
PF 27-APR-1999; 99WO-CA00375.
XX
XX
PR 27-APR-1998; 98CA-2230201.
XX
XX 05-FEB-1999; 99US-0118739.
XX
XX
PA (HSCR-) HSC RES & DEV LP.
XX
XX
PI Egan SE, Wang W, Sengar A;
XX
XX WPI; 2000-052802/04.

DR P-PSDB; AAY57444.
 XX New nucleic acid encoding Ese1 and 2 proteins, involved in regulation
 PT of endocytosis, used e.g. for treating cancer or preventing viral
 PT infection
 XX
 PS Claim 6; Page 40-42; 9pp; English.
 XX
 CC The present sequence encodes mouse Ese1. The present invention
 CC specifically describes mammalian Ese1 and 2 proteins (1) and their splice
 CC variants (Ese = EH-domain and SH3-domain regulator of endocytosis). (1)
 CC are involved in regulation of clathrin-mediated endocytosis (as a complex
 CC with Eps15 protein), vesicular trafficking and actin cytoskeleton.
 CC Generally (1) or its (ant)agonists, mimetics, fragments and inactive
 CC mutants); (1)-specific antibodies (Ab); sequences antisense to the (1)
 CC polynucleotide; agents that downregulate expression of Ese genes or
 CC antagonists of an Ese binding partner are used to treat diseases
 CC associated with undesirable endocytosis and resulting changes in cellular
 CC function. Particularly overexpression of Ese1 is used to block
 CC clathrin-mediated endocytosis in vivo or in cell cultures, while
 CC administration of (1) is used to promote endocytosis of selected cells.
 CC (Ant)agonists of (1) or Ab are used to suppress abnormal proliferation of
 CC cells that can be stimulated to proliferate by a growth factor receptor;
 CC and similar compounds (also inactive Ese mutants) can be used to prevent
 CC viral infection. Endocytosis may also be regulated, in vivo or in cell
 CC cultures, by forming an Ese-Eps15 complex, then binding dynamin to the
 CC complex. Generally conditions that can be treated include cancer;
 CC abnormal cell division or migration; viral infection; or abnormal
 CC receptor signalling, tissue development or synaptic transmission.
 CC
 SQ Sequence 3723 BP; 1065 A; 941 C; 1025 G; 692 T; 0 other;

Query Match 100.0%; Score 3642; DB 21; Length 3723;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 3642; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCTGAGTTTCCACACCTTTCGGTGGTACCTGATGTCCTGGGCCATTAAGTGGAG 60
 Db 1 ATGGCTGAGTTTCCACACCTTTCGGTGGTACCTGATGTCCTGGGCCATTAAGTGGAG 60
 QY 61 GAAAGGGCCAAAGCATGACAGCAGAGTCTTACCTGAAAGCCGATGACGGGATTTATTA 120
 Db 61 GAAAGGGCCAAAGCATGACAGCAGAGTCTTACCTGAAAGCCGATGACGGGATTTATTA 120
 QY 121 GGTATTCAGAGGAGAACTTTTTCATCTGGGTTACTGACCTGCTTTAGCAACA 180
 Db 121 GGTATTCAGAGGAGAACTTTTTCATCTGGGTTACTGACCTGCTTTAGCAACA 180
 QY 181 ATATGGGGCTAGGAGCATGAATAAGATGGAAGATGATCAATGGAATTTTCCATA 240
 Db 181 ATATGGGGCTAGGAGCATGAATAAGATGGAAGATGATCAATGGAATTTTCCATA 240
 QY 241 GCCATGAAGCTTATCAAACTGAAGCTAAGGATATCACTCCCTCCACAATTCCTCCCT 300
 Db 241 GCCATGAAGCTTATCAAACTGAAGCTAAGGATATCACTCCCTCCACAATTCCTCCCT 300
 QY 301 GTCATGAAGCAAGCAACAGTGGCTATTTCCAGTGCACACAGCATTTGGTATAGAGGAT 360
 Db 301 GTCATGAAGCAAGCAACAGTGGCTATTTCCAGTGCACACAGCATTTGGTATAGAGGAT 360
 QY 361 GCTACATGACCACTACAGCTGTGCTCTGTCGCAATGGGCTCCATTCAGATTGT 420
 Db 361 GCTACATGACCACTACAGCTGTGCTCTGTCGCAATGGGCTCCATTCAGATTGT 420
 QY 421 GGAATGTCACACCTTAGTATCTTCTGCTCCAGAGCAGTGCCTCCCTGGCTAAC 480
 Db 421 GGAATGTCACACCTTAGTATCTTCTGCTCCAGAGCAGTGCCTCCCTGGCTAAC 480
 QY 481 GGGGCTCTCCCGTATACAGCTTGTGCTGGTGGGATCTGTCAGCCACATGGCCA 540
 Db 481 GGGGCTCTCCCGTATACAGCTTGTGCTGGTGGGATCTGTCAGCCACATGGCCA 540
 QY 541 AAGAGTTCTTCTTACAGAGATCTGGTCCAGGGTCAACAATTAAACATTAAGTTACGAG 600

Db 541 AAGAGTTCTTCTTACAGAGATCTGGTCCAGGGTCAACAATTAAACATTAAGTTACGAG 600
 QY 601 GCAACATCATTTGATGTCGCCAGCGCCCTTCACAGCAGAAATGGCTGTGCTCACTCA 660
 Db 601 GCAACATCATTTGATGTCGCCAGCGCCCTTCACAGCAGAAATGGCTGTGCTCACTCA 660
 QY 661 TCAAGGCTGAAATACAGGAGTTTATCAAGCCACGACAAATATGAGTGAACACTTA 720
 Db 661 TCAAGGCTGAAATACAGGAGTTTATCAAGCCACGACAAATATGAGTGAACACTTA 720
 QY 721 ACAGGTCGCCAGGCAAGAACTATTTCATGCAATCAAGTTTATCCAGGCTCACTGAGCT 780
 Db 721 ACAGGTCGCCAGGCAAGAACTATTTCATGCAATCAAGTTTATCCAGGCTCACTGAGCT 780
 QY 781 TCAATATGAAATCTTTTGCATTTGATCAAGATGAGAAAATCACTGCAAGAAATTTATC 840
 Db 781 TCAATATGAAATCTTTTGCATTTGATCAAGATGAGAAAATCACTGCAAGAAATTTATC 840
 QY 841 CTAGCTATGCACTTAATTTGATGTTGCAATGTCGTCAGCAGTCCGCGCTGCTGCT 900
 Db 841 CTAGCTATGCACTTAATTTGATGTTGCAATGTCGTCAGCAGTCCGCGCTGCTGCTGCT 900
 QY 901 CCAAGATATACATCCCTCTTCTCAGAAAGTTGCTCCGCGCAGTGGAGTCCGTCATA 960
 Db 901 CCAAGATATACATCCCTCTTCTCAGAAAGTTGCTCCGCGCAGTGGAGTCCGTCATA 960
 QY 961 AGCTCTTCTTGTGATTCAGAGGCTGCTGAGAGCCGCTGTCAGAGATGACAGCAG 1020
 Db 961 AGCTCTTCTTGTGATTCAGAGGCTGCTGAGAGCCGCTGTCAGAGATGACAGCAG 1020
 QY 1021 CCAGAGAAAGAACTGCTGTCATTTGAAGATGAAGACGGGGAATCTTGACAGAGGC 1080
 Db 1021 CCAGAGAAAGAACTGCTGTCATTTGAAGATGAAGACGGGGAATCTTGACAGAGGC 1080
 QY 1081 AGTGTGAGCTGAGAAAGCCGCGCAAGCGCTTTGAGAGCAGCAGCAAGACAGAG 1140
 Db 1081 AGTGTGAGCTGAGAAAGCCGCGCAAGCGCTTTGAGAGCAGCAGCAAGACAGAG 1140
 QY 1141 CGGTTGGCTCAGCTGAGAGCGCGCGAGCAGAGAGAAAGCCGGAGCGCAGAGCAG 1200
 Db 1141 CGGTTGGCTCAGCTGAGAGCGCGCGAGCAGAGAGAAAGCCGGAGCGCAGAGCAG 1200
 QY 1201 GAGGCGCAAGCGGCGCTGAGAGCTGGAAGAGCTGGAAGAGCAGCGGAGCTGAGAGCG 1260
 Db 1201 GAGGCGCAAGCGGCGCTGAGAGCTGGAAGAGCTGGAAGAGCAGCGGAGCTGAGAGCG 1260
 QY 1261 CAGCGAGAGAGAGAGAGAGAGATCGAGAGCGCGAGCGCGCAAAACGGGAACTG 1320
 Db 1261 CAGCGAGAGAGAGAGAGAGAGATCGAGAGCGCGAGCGCGCGCAAAACGGGAACTG 1320
 QY 1321 GAAGAGCAGCGCAACTTGAATGGGAAACGGAACCGAGACAGAACTCTGTAATCAGAGG 1380
 Db 1321 GAAGAGCAGCGCAACTTGAATGGGAAACGGAACCGAGACAGAACTCTGTAATCAGAGG 1380
 QY 1381 AACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
 Db 1381 AACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
 QY 1441 TTGAAGCTCTGTAATGCAAAAAGCATCAGCTAGAGAGAAATCTTACAGATATCAGGTG 1500
 Db 1441 TTGAAGCTCTGTAATGCAAAAAGCATCAGCTAGAGAGAAATCTTACAGATATCAGGTG 1500
 QY 1501 CGACTGGCAACCCAGAGGCAAGAAATTTGAGAGCAGAAAGTCTAGAGAGCTTAAGATT 1560
 Db 1501 CGACTGGCAACCCAGAGGCAAGAAATTTGAGAGCAGAAAGTCTAGAGAGCTTAAGATT 1560
 QY 1561 GCTGAATTCACCCCACTTAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620
 Db 1561 GCTGAATTCACCCCACTTAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620
 QY 1621 ATTCCAGAGAAACAGATCTCAGTGAACAGTTAAACAGATCCAGCAAGAGTTTCAT 1680

Db 1621 ATTCAGAGAAACAGATCTAGTGAACAGTTAAACAGTCCAGAGAAAGATTTCAT 1680
 QY 1681 AAGAGATCGCTTCTTACCTCAAAAAGAGCTTGAAGCAAGAGAGTGGCCGGCAGCAG 1740
 Db 1681 AAGAGATCGCTTCTTACCTCAAAAAGAGCTTGAAGCAAGAGAGTGGCCGGCAGCAG 1740
 QY 1741 CTCGGGAGCAGCTGAGCAGAGGTGAGAGAGAACAGGTCAAGTGTGACAGAGATTGAT 1800
 Db 1741 CTCGGGAGCAGCTGAGCAGAGGTGAGAGAGAACAGGTCAAGTGTGACAGAGATTGAT 1800
 QY 1801 GTTTTCAACACCAAGTGAAGAACTGAGAGATTCATGCAACAGCAATCTCCAGAG 1860
 Db 1801 GTTTTCAACACCAAGTGAAGAACTGAGAGATTCATGCAACAGCAATCTCCAGAG 1860
 QY 1861 CAGAGTCCCTGGAGCAGCGGAGCTGAAGCAAGAGAGAGAGAGAGAGAGAGAGAGAG 1920
 Db 1861 CAGAGTCCCTGGAGCAGCGGAGCTGAAGCAAGAGAGAGAGAGAGAGAGAGAGAGAG 1920
 QY 1921 TTAGAGAGCAAAAGAGAGAGAGCTCAGAGAGAGATTCAGGAAAGGAGCAAGATGGCTG 1980
 Db 1921 TTAGAGAGCAAAAGAGAGAGAGCTCAGAGAGAGATTCAGGAAAGGAGCAAGATGGCTG 1980
 QY 1981 GAGCATGTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2040
 Db 1981 GAGCATGTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2040
 QY 2041 AAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2100
 Db 2041 AAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2100
 QY 2101 GACAAGCAGAGTGGCTTTTCCATCCGATCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2160
 Db 2101 GACAAGCAGAGTGGCTTTTCCATCCGATCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2160
 QY 2161 CCTGTGCTTACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2220
 Db 2161 CCTGTGCTTACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2220
 QY 2221 GATATATACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2280
 Db 2221 GATATATACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2280
 QY 2281 CCAGAGATATATAGTCACTGATGATGAAGCAGAGCTGAGAGAGAGAGAGAGAGAGAG 2340
 Db 2281 CCAGAGATATATAGTCACTGATGATGAAGCAGAGCTGAGAGAGAGAGAGAGAGAGAG 2340
 QY 2341 GAGCTGAAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2400
 Db 2341 GAGCTGAAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2400
 QY 2401 GAGGTTCCCATCTCAGCCCAAAACAGATGACCATCTGCCCCCTGCCCCCAAACTG 2460
 Db 2401 GAGGTTCCCATCTCAGCCCAAAACAGATGACCATCTGCCCCCTGCCCCCAAACTG 2460
 QY 2461 GCTCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2520
 Db 2461 GCTCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2520
 QY 2521 AACCACTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2580
 Db 2521 AACCACTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2580
 QY 2581 GACCACTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2640
 Db 2581 GACCACTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2640
 QY 2641 CGGCAAGATCAGCTTTTACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2700
 Db 2641 CGGCAAGATCAGCTTTTACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2700
 QY 2701 GGCAGAGGTGAAGAGGTGAAGAGGTGAAGAGGTGAAGAGGTGAAGAGGTGAAGAGGTGA 2760
 Db 2701 GGCAGAGGTGAAGAGGTGAAGAGGTGAAGAGGTGAAGAGGTGAAGAGGTGAAGAGGTGA 2760

QY 2761 AAAAGCAACCACTTAATTTTAAACAAAGGTGACGTATCACCCTTCTGGAACAGCAAGAC 2820
 Db 2761 AAAAGCAACCACTTAATTTTAAACAAAGGTGACGTATCACCCTTCTGGAACAGCAAGAC 2820
 QY 2821 ATGTGGTGTGTTGAGAAAGTTCAAGGTCAAGAGGTTGGTTCCTCAAGTCTTAAGTGA 2880
 Db 2821 ATGTGGTGTGTTGAGAAAGTTCAAGGTCAAGAGGTTGGTTCCTCAAGTCTTAAGTGA 2880
 QY 2881 CTCAATTCAGGGCCCGTAAAGAAATTCACAGATTCGATTCGGCCCTTAAGTGAAGTCT 2940
 Db 2881 CTCAATTCAGGGCCCGTAAAGAAATTCACAGATTCGATTCGGCCCTTAAGTGAAGTCT 2940
 QY 2941 GCTAGCTTAAAGAGAGTGGCTTCCCGGCGGCAAGCCAGCATTCGCGGAGAGAGATT 3000
 Db 2941 GCTAGCTTAAAGAGAGTGGCTTCCCGGCGGCAAGCCAGCATTCGCGGAGAGAGATT 3000
 QY 3001 ATTGCAATGTACACATACAGAGATTCTGAGCAAGAGATTTAACTTTTCAGCAAGGGAGT 3060
 Db 3001 ATTGCAATGTACACATACAGAGATTCTGAGCAAGAGATTTAACTTTTCAGCAAGGGAGT 3060
 QY 3061 GTGATTTGTGTTTACCAAGAAAGTGTGACTGTGAGCGGAAACGGTGGGCGACAAAGTCC 3120
 Db 3061 GTGATTTGTGTTTACCAAGAAAGTGTGACTGTGAGCGGAAACGGTGGGCGACAAAGTCC 3120
 QY 3121 GGAAGTCTTCCCTTACACTGATGAGAGCTTAAGATTCAGAGAGGCTTGGAACTGCTGGG 3180
 Db 3121 GGAAGTCTTCCCTTACACTGATGAGAGCTTAAGATTCAGAGAGGCTTGGAACTGCTGGG 3180
 QY 3181 AAAACAGGAGTTTGAAGAAAGAAACCTGAATTTGCCAGGTTATTGCTTCTACGCTGCT 3240
 Db 3181 AAAACAGGAGTTTGAAGAAAGAAACCTGAATTTGCCAGGTTATTGCTTCTACGCTGCT 3240
 QY 3241 ACTGTGCTCCGAAACAACCTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3300
 Db 3241 ACTGTGCTCCGAAACAACCTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3300
 QY 3301 CCAGGTGATGAGTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3360
 Db 3301 CCAGGTGATGAGTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3360
 QY 3361 TTTTCAAGCAATTAATGTCAAACTTTAAGCCCGGAGCAAGCAAAATCAACCCCACTGAG 3420
 Db 3361 TTTTCAAGCAATTAATGTCAAACTTTAAGCCCGGAGCAAGCAAAATCAACCCCACTGAG 3420
 QY 3421 CTACCCAGAGCCGAGTGCAGCAGCAGTGTGCTCAGGTGATCGGAGTGAAGATTACACC 3480
 Db 3421 CTACCCAGAGCCGAGTGCAGCAGCAGTGTGCTCAGGTGATCGGAGTGAAGATTACACC 3480
 QY 3481 GCCCAAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3540
 Db 3481 GCCCAAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3540
 QY 3541 GAGGAGCCGGAATGTGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3600
 Db 3541 GAGGAGCCGGAATGTGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3600
 QY 3601 TATGTAAAGCTGACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3642
 Db 3601 TATGTAAAGCTGACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3642

RESULT 2
 AA239025
 ID AA239025 standard; cDNA; 5144 BP.
 XX
 AC AA239025;
 XX
 DT 28-FEB-2000 (first entry)
 XX
 DE Mouse Bse1 coding sequence.
 XX
 KW Mouse; murine; Bse1; Bse2; endocytosis; vesicular trafficking;

Db 1321 GAAAGGACGACAACTTGAATGGGAAACGGAAACGGAGACAGGAACCTCTGAATCAGAGG 1380
Qy 1381 AACAAAGACAGAGGAGGACCGGTGTCCTGAAGGCAAGAGAGAGACTCTGAGTTTGAG 1440
Db 1381 AACAAAGACAGAGGAGGACCGGTGTCCTGAAGGCAAGAGAGAGACTCTGAGTTTGAG 1440
Qy 1441 TTGAGAGCTCTGAATGACAAAAGCATCAGTGAAGAAAATTGAGATATCAGGTGT 1500
Db 1441 TTGAGAGCTCTGAATGACAAAAGCATCAGTGAAGAAAATTGAGATATCAGGTGT 1500
Qy 1501 CGACTGGCAACCCAGAGGCAAGAAATTGAGAGCAAGCAAGCTTGAAGAGCTTAAGAAAT 1560
Db 1501 CGACTGGCAACCCAGAGGCAAGAAATTGAGAGCAAGCAAGCTTGAAGAGCTTAAGAAAT 1560
Qy 1561 GGTGAATACCCACTTACAGAGCAAGTTGAGAGAAATCTCAGAAATGCTTGAAGACTT 1620
Db 1561 GGTGAATACCCACTTACAGAGCAAGTTGAGAGAAATCTCAGAAATGCTTGAAGACTT 1620
Qy 1621 ATTCCAGAGAAACAGATACCTCAGTGAACAGTTAAAAGTCCAGAGAAACAGTTGAT 1680
Db 1621 ATTCCAGAGAAACAGATACCTCAGTGAACAGTTAAAAGTCCAGAGAAACAGTTGAT 1680
Qy 1681 AAGAGCTCGCTTCTTACCTCAAAGAGCTTGAAGCAAGAGAGCTGGCCGGCAGCAG 1740
Db 1681 AAGAGCTCGCTTCTTACCTCAAAGAGCTTGAAGCAAGAGAGCTGGCCGGCAGCAG 1740
Qy 1741 CTCCGGAGACAGCTGAGACAGGTGAGAGAGAGCAAGTCAAGTCAAGTCTGAGAGATGAT 1800
Db 1741 CTCCGGAGACAGCTGAGACAGGTGAGAGAGAGCAAGTCAAGTCAAGTCTGAGAGATGAT 1800
Qy 1801 GTTTTCAACAAACAGCTGAAGAACTGAGAGAGATCATAGCAAAACAGCAATCCAGAG 1860
Db 1801 GTTTTCAACAAACAGCTGAAGAACTGAGAGAGATCATAGCAAAACAGCAATCCAGAG 1860
Qy 1861 CAGAGTCCCTGAGAGCAGCGGACTGAAGCAAGAAAGACAGAGAGAGAGAGAGCTGAG 1920
Db 1861 CAGAGTCCCTGAGAGCAGCGGACTGAAGCAAGAAAGACAGAGAGAGAGAGAGCTGAG 1920
Qy 1921 TTGAGAGCAAAAGAGAAACCTCAGAGACAGATTCAGAGAAAGGAGCAAGCAATGGCTG 1980
Db 1921 TTGAGAGCAAAAGAGAAACCTCAGAGACAGATTCAGAGAAAGGAGCAAGCAATGGCTG 1980
Qy 1981 GAGCATGTGACAGAGAGAGAGCAGCAGCCCGGAGAAACCCCAAGAGAGAGAGAGACTG 2040
Db 1981 GAGCATGTGACAGAGAGAGAGCAGCAGCCCGGAGAAACCCCAAGAGAGAGAGAGACTG 2040
Qy 2041 AAGAGGAAAGACAGTGTCAAGAAAGAGAGGCGGAAAGAGAGAGAGAGAGAGAGAGAG 2100
Db 2041 AAGAGGAAAGACAGTGTCAAGAAAGAGAGGCGGAAAGAGAGAGAGAGAGAGAGAGAG 2100
Qy 2101 GACAAAGCAAGTCGGCTTTTTCATCGCATCAGAGAGCAGCTAAGCTGCGACCCAGGGA 2160
Db 2101 GACAAAGCAAGTCGGCTTTTTCATCGCATCAGAGAGCAGCTAAGCTGCGACCCAGGGA 2160
Qy 2161 CCTGGTCTACCAAGAGAAAGGCGCTTACCATTTTGCACAGAGAGGTGAAAGAGT 2220
Db 2161 CCTGGTCTACCAAGAGAAAGGCGCTTACCATTTTGCACAGAGAGGTGAAAGAGT 2220
Qy 2221 GTATATTACCGAGCGCTGTACCTCTTGAATCAGAAAGTCAAGATGATCAACATCCAG 2280
Db 2221 GTATATTACCGAGCGCTGTACCTCTTGAATCAGAAAGTCAAGATGATCAACATCCAG 2280
Qy 2281 CAGAGAGATATGATCTATGTGTGATGAAGACAGAGCTGAGAGAGAGAGAGAGAGAGAG 2340
Db 2281 CAGAGAGATATGATCTATGTGTGATGAAGACAGAGCTGAGAGAGAGAGAGAGAGAGAG 2340
Qy 2341 GAGCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2400
Db 2341 GAGCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2400
Qy 2401 GAGGTTCCACTTCAGCAGCAAAACAGTGAACGATCTGAGATCTGAGAGAGAGAGAGAG 2460
Db 2401 GAGGTTCCACTTCAGCAGCAAAACAGTGAACGATCTGAGATCTGAGAGAGAGAGAGAGAG 2460

Qy 2461 GCTTCGGTGAAGACCCCTGCTCTTTGACAGTGAACCTTCTGAGAGCCCTCCACAAACCCC 2520
Db 2461 GCTTCGGTGAAGACCCCTGCTCTTTGACAGTGAACCTTCTGAGAGCCCTCCACAAACCCC 2520
Qy 2521 AACAACTGGGAGACTTCACTGAGTCCAGTGGCCAGAGAGAGAGAGAGAGAGAGAGAGAG 2580
Db 2521 AACAACTGGGAGACTTCACTGAGTCCAGTGGCCAGAGAGAGAGAGAGAGAGAGAGAGAG 2580
Qy 2581 GACAACTGGGAGACTTCACTGAGTCCAGTGGCCAGAGAGAGAGAGAGAGAGAGAGAGAG 2640
Db 2581 GACAACTGGGAGACTTCACTGAGTCCAGTGGCCAGAGAGAGAGAGAGAGAGAGAGAGAG 2640
Qy 2641 CGGAGAGATCAGAGCTTTTACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2700
Db 2641 CGGAGAGATCAGAGCTTTTACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2700
Qy 2701 GGCAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2760
Db 2701 GGCAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2760
Qy 2761 AAAGCAACCACTTAATTTTAAACAAAGTGAAGTATCAGCTTCTGAGAGAGAGAGAGAG 2820
Db 2761 AAAGCAACCACTTAATTTTAAACAAAGTGAAGTATCAGCTTCTGAGAGAGAGAGAGAG 2820
Qy 2821 ATGTGTGTGTGTGAGAGAGTCAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2880
Db 2821 ATGTGTGTGTGTGAGAGAGTCAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2880
Qy 2881 CTCAATTCAGAGGCGCGTGAAGAAATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2940
Db 2881 CTCAATTCAGAGGCGCGTGAAGAAATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2940
Qy 2941 GCTAGTCTTAAAGAGAGTGTTCCTCCCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAG 3000
Db 2941 GCTAGTCTTAAAGAGAGTGTTCCTCCCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAG 3000
Qy 3001 ATTGCAATGTACATCAGAGAGTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3060
Db 3001 ATTGCAATGTACATCAGAGAGTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3060
Qy 3061 GTGATTTGTGTTTAAACAAAGAGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3120
Db 3061 GTGATTTGTGTTTAAACAAAGAGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3120
Qy 3121 GAGAGTCTTCCCTTCACTATGTAGAGCTTAAAGATTCAGAGAGAGAGAGAGAGAGAGAG 3180
Db 3121 GAGAGTCTTCCCTTCACTATGTAGAGCTTAAAGATTCAGAGAGAGAGAGAGAGAGAGAG 3180
Qy 3181 AAAAGAGAGAGTTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3240
Db 3181 AAAAGAGAGAGTTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3240
Qy 3241 ACTGATCCCGAACAACCTCAGCAGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3300
Db 3241 ACTGATCCCGAACAACCTCAGCAGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3300
Qy 3301 CAGAGTGTGATGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3360
Db 3301 CAGAGTGTGATGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3360
Qy 3361 TTTTCAAGCAAAATTAATGTAATCTTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3420
Db 3361 TTTTCAAGCAAAATTAATGTAATCTTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3420
Qy 3421 CTACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3480
Db 3421 CTACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3480
Qy 3481 GGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3540
Db 3481 GGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3540

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1

Db 1280 CCAGAGAGAACTGCTGTGACATTGAAGATAGAGCGGAGAACTTGACGAGGC 1339
QY 1081 AGTGTGAGCTGAGAGAGCGCGCCAAAGCGCTCTTGAGAGAGAGCGAAAGAGAGAG 1140
Db 1340 AGGTGTGAGCTGAGAGAGCGCGCCAAAGCGCTCTTGAGAGAGAGCGAAAGAGAGAG 1399
QY 1141 CGGTGTGCTGAGTGAAGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
Db 1400 CGGTGTGCTGAGTGAAGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1459
QY 1201 GAGGCCAAGCGGAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
Db 1460 GAGGCCAAGCGGAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1519
QY 1261 GAGGCCAAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
Db 1520 GAGGCCAAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1579
QY 1321 GAAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
Db 1580 GAAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1639
QY 1381 AACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
Db 1640 AACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1699
QY 1441 TTAGAAGCTCTGAATGACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500
Db 1700 TTAGAAGCTCTGAATGACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1759
QY 1501 CGAGTGGCAACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560
Db 1760 CGAGTGGCAACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1819
QY 1561 GCTGAAGATCAACCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620
Db 1820 GCTGAAGATCAACCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1879
QY 1621 ATTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1680
Db 1880 ATTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1939
QY 1681 AGAGACTGCTCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1740
Db 1940 AGAGACTGCTCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1999
QY 1741 CTCGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1800
Db 2000 CTCGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2059
QY 1801 GTTTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860
Db 2060 GTTTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2119
QY 1861 CAGAGGTCCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1920
Db 2120 CAGAGGTCCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2179
QY 1921 TTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1980
Db 2180 TTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2239
QY 1981 GAGCAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2040
Db 2240 GAGCAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2299
QY 2041 AAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2100
Db 2300 AAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2359
QY 2101 GACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2160
Db 2360 GACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2419

QY 2161 CCCTGGTCTACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2220
Db 2420 CCCTGGTCTACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2479
QY 2221 GTATATTTACGAGCGGTGTACCTTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2280
Db 2480 GTATATTTACGAGCGGTGTACCTTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2539
QY 2281 CCAAGAGATATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2340
Db 2540 CCAAGAGATATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2599
QY 2341 GAGCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2400
Db 2600 GAGCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2659
QY 2401 GAGGTTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCT 2460
Db 2660 GAGGTTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCT 2719
QY 2461 GCTCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2520
Db 2720 GCTCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2779
QY 2521 AACAACTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2580
Db 2780 AACAACTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2839
QY 2581 GACAACTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2640
Db 2840 GACAACTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2899
QY 2641 CGGAGAGATCAAGCTTTTACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2700
Db 2900 CGGAGAGATCAAGCTTTTACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2959
QY 2701 GGGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2760
Db 2960 GGGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3019
QY 2761 AAGAGCAACCACTTAAATTTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2820
Db 3020 AAGAGCAACCACTTAAATTTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3079
QY 2821 ATGTGTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2880
Db 3080 ATGTGTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3139
QY 2881 CTCATTTTCAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2940
Db 3140 CTCATTTTCAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3199
QY 2941 GCTAGCTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3000
Db 3200 GCTAGCTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3259
QY 3001 ATTGCTATGTACATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3060
Db 3260 ATTGCTATGTACATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3319
QY 3061 GTGATTTGTGTGTACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3120
Db 3320 GTGATTTGTGTGTACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3379
QY 3121 GAGAGCTTCCCTTCTTAACTATGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3180
Db 3380 GAGAGCTTCCCTTCTTAACTATGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3439
QY 3181 AAAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3240
Db 3440 AAAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3499


```

Oy 3241 ACTGTCGCCGAACAACCTCAGCCCTGCTCTGGAGAGCTGATTTCTGATCCGGAAAAAGAAC 3300
Db 3500 ACTGATCCCGAACAACCTCAGCCCTGCTCTGGAGAGCTGATTTCTGATCCGGAAAAAGAAC 3559
Oy 3301 CCAGGTGATGATGGGGAAGAGAACTGCAAGCTCGAGGGAAAAAGCCGAGATAGGGTGG 3360
Db 3560 CCAGGTGATGATGGGGAAGAGAACTGCAAGCTCGAGGGAAAAAGCCGAGATAGGGTGG 3619
Oy 3361 TTTCCAGCAAAATTAATGTAATCTTTAGCCCCCGGAAACCAACCAAAATACCCCAACTAG 3420
Db 3620 TTTCCAGCAAAATTAATGTAATCTTTAGCCCCCGGAAACCAACCAAAATACCCCAACTAG 3679
Oy 3421 CTACCCAGAACCGAGCTGAGCAGCAGCAGTGTGCTCAGTGTATGCGGATGATGATTAACAC 3480
Db 3680 CTACCCAGAACCGAGCTGAGCAGCAGCAGTGTGCTCAGTGTATGCGGATGATGATTAACAC 3739
Oy 3481 GCCCAGAACGATGACGAACCTAGCCTTCAGCAAAAGCCAGATCATCAACGCTCTCAACAG 3540
Db 3740 GCCCAGAACGATGACGAACCTAGCCTTCAGCAAAAGCCAGATCATCAACGCTCTCAACAG 3799
Oy 3541 GAGAACCCGAGCTGATGGAAGAGAAAGTCAATGAGGCTCTTCCCATCCAT 3600
Db 3800 GAGAACCCGAGCTGATGGAAGAGAAAGTCAATGAGGCTCTTCCCATCCAT 3859
Oy 3601 TATGTAAGCTGACCAACAGATGAGACCCAGCCAGCAATG 3641
Db 3860 TATGTAAGCTGACCAACAGATGAGACCCAGCCAGCAATG 3900

```

```

RESULT 4
AAZ3908
ID AAZ3908 standard; cDNA; 5082 BP.
AC AAZ3908;
XX
XX 28-FEB-2000 (first entry)
DT
XX
XX Mouse Eae1 full length cDNA sequence.
DE
XX
XX Mouse; murine; Eae1; Eae2; endocytosis; vesicular trafficking;
KM regulation; actin cytoskeleton; detection; cancer; infection;
KM EH-domain and SH3-domain regulator of endocytosis; anticancer;
KM antiproliferative; antiviral; ss.
XX
XX Mus sp.
OS
XX
XX W09955728-A2.
PN
XX
XX 04-NOV-1999.
PD
XX
XX 27-APR-1999; 99MO-CA00375.
PF
XX
XX 27-APR-1998; 98CA-2230201.
PR
XX 05-FEB-1999; 99US-0118739.
XX
XX (HSCR-) HSC RES & DEV LP.
PA
XX
XX Egan SE, Wang W, Sengar A;
PI
XX
XX WPI; 2000-052802/04.
DR
XX
XX P-PSDB; AAY57444.
PT
XX
XX New nucleic acid encoding Eae1 and 2 proteins, involved in regulation
PT of endocytosis, used e.g. for treating cancer or preventing viral
PT infection.
XX
XX
XX Claim 6; Page 38-40; 99pp; English..
PS
XX
XX The present sequence encodes mouse Eae1. The present invention
CC specifically describes mammalian Eae1 and 2 proteins (1) and their splice
CC variants (Ese = EH-domain and SH3-domain regulator of endocytosis). (1)
CC are involved in regulation of clathrin-mediated endocytosis (as a complex
CC with Eap1s protein), vesicular trafficking and actin cytoskeleton.

```

Generally (1) (or its (ant)agonists, mimetics, fragments and inactive mutants); (1)-specific antibodies (Ab); sequences antisense to the (1) polynucleotide; agents that downregulate expression of Eae genes or the (ant)agonists of an Eae binding partner are used to treat diseases associated with undesirable endocytosis and resulting changes in cellular function. Particularly overexpression of Eae1 is used to block clathrin-mediated endocytosis in vivo or in cell cultures, while administration of (1) is used to promote endocytosis of selected cells. (Ant)agonists of (1) or Ab are used to suppress abnormal proliferation of cells that can be stimulated to proliferate by a growth factor receptor; and similar compounds (also inactive Eae mutants) can be used to prevent viral infection. Endocytosis may also be regulated, in vivo or in cell cultures, by forming an Eae-Esp1s complex, then binding dynamin to the complex. Generally conditions that can be treated include cancer; abnormal cell division or migration; viral infection; or abnormal receptor signalling, tissue development or synaptic transmission.

Sequence 5082 BP; 1363 A; 1253 C; 1391 G; 1075 T; 0 other;

Query Match 99.7%; Score 3630; DB 21; Length 5082;

Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Indels 1; Gaps 1;

Matches 3641; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

```

Oy 1 ATGAGCTCAGTTTCCACACCTTTGGTGTAGCCTGATGTCGTGGCCATTACTGTGAG 60
Db 259 ATGAGCTCAGTTTCCACACCTTTGGTGTAGCCTGATGTCGTGGCCATTACTGTGAG 318
Oy 61 GAAAGGGCCCAAGCACTGACCAAGCACTTCTTACCTGAAAGCCGATAGGGGATTTATCT 120
Db 319 GAAAGGGCCCAAGCACTGACCAAGCACTTCTTACCTGAAAGCCGATAGGGGATTTATCT 378
Oy 121 GGTATCAAGGAGAGAACTTTTTCATCTGGGTTTACCTGAGCTGCTTGTAGCAAA 180
Db 379 GGTATCAAGGAGAGAACTTTTTCATCTGGGTTTACCTGAGCTGCTTGTAGCAAA 438
Oy 181 ATATGGCGCTAGGCGACATGAATACATGAGAGATGATGATCAAGTGAATTTTCATA 240
Db 439 ATATGGCGCTAGGCGACATGAATACATGAGAGATGATGATCAAGTGAATTTTCATA 498
Oy 241 GCCATGAAGCTTATCAACTGAAGCTACAGAGATATAGTCCCTCCACACTTCCCTCCT 300
Db 499 GCCATGAAGCTTATCAACTGAAGCTACAGAGATATAGTCCCTCCACACTTCCCTCCT 558
Oy 301 GTCATGAACAGCAACAGAGTGGCTATTTCAGTGCACAGCATTTGGTATAGAGGGATT 360
Db 559 GTCATGAACAGCAACAGAGTGGCTATTTCAGTGCACAGCATTTGGTATAGAGGGATT 618
Oy 361 GCTAGCATGCCACCACTCACAGCTGTGCTCTGTCGCAATGGGCTCATTCAGTTGTT 420
Db 619 GCTAGCATGCCACCACTCACAGCTGTGCTCTGTCGCAATGGGCTCATTCAGTTGTT 678
Oy 421 GGAATGTCTCCACCTTATATCTTCTGTCCTCCACAGAGAGTCCCTCCCTGGCTAAC 480
Db 679 GGAATGTCTCCACCTTATATCTTCTGTCCTCCACAGAGAGTCCCTCCCTGGCTAAC 738
Oy 481 GGGGCTCTCCGCTATACAGCCTCTGCTGCTGCTTGGAGATCCTTGAGCCACATGGCCA 540
Db 739 GGGGCTCTCCGCTATACAGCCTCTGCTGCTGCTTGGAGATCCTTGAGCCACATGGCCA 798
Oy 541 AAGAGTTCTTCTTCCAGCAGATCTGATCAGGGTCAACATTTAAACATAAGTTACAGAG 600
Db 799 AAGAGTTCTTCTTCCAGCAGATCTGATCAGGGTCAACATTTAAACATAAGTTACAGAG 858
Oy 601 GCAACATCATTCATGTCGCGAGGCCCTTCACAGCAGAGAAATGGCTGCTCCACAGTCA 660
Db 859 GCAACATCATTCATGTCGCGAGGCCCTTCACAGCAGAGAAATGGCTGCTCCACAGTCA 918
Oy 661 TCAAGGCTGAATATCAGCAGATTTATCAACAGCCAGCAAAATCATAGATGAGCACTTA 720
Db 919 TCAAGGCTGAATATCAGCAGATTTATCAACAGCCAGCAAAATCATAGATGAGCACTTA 978
Oy 721 ACAGGTCGCCAGGAGAACTATTCTCATGCAATCAAGTTTACCCAGGCTCAAGCTGGCT 780

```

Db 979 AAGAGTCCCAAGCAAGAACTATTCTCATGATCAATCAAGTTTACCAGAGCTCAGTGGCT 1038
QY 781 TCAATATGGAATCTTCTGACATTTGATCAAGATGAGAAATCTGACAGAAATTTATC 840
Db 1039 TCAATATGGAATCTTCTGACATTTGATCAAGATGAGAAATCTGACAGAAATTTATC 1098
QY 841 CTAGCTATGACACCTAAATGATGTTGGCATGTCTGTACAGCACTGCGCCGCTCTGCT 900
Db 1099 CTAGCTATGACACCTAAATGATGTTGGCATGTCTGTACAGCACTGCGCCGCTCTGCT 1158
QY 901 CCAAGATACATCCCTCTCTCTCTCAGAAAGATTCCTCCGAGCTGAGATGTCCTCAT 960
Db 1159 CCAAGATACATCCCTCTCTCTCTCAGAAAGATTCCTCCGAGCTGAGATGTCCTCAT 1218
QY 961 AGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1020
Db 1219 AGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1278
QY 1021 CCAGAGAAAGAACTGCTCTGATCAATTTGAAGATPAAGAGCGGAGAACTTCCAGCGAGC 1080
Db 1279 CCAGAGAAAGAACTGCTCTGATCAATTTGAAGATPAAGAGCGGAGAACTTCCAGCGAGC 1338
QY 1081 AGTGTGAGCTGAGAAAGCGCGCCCAAGCGCTCTTGAAGCAGCGCAAGAGCAGAG 1140
Db 1339 AGTGTGAGCTGAGAAAGCGCGCCCAAGCGCTCTTGAAGCAGCGCAAGAGCAGAG 1398
QY 1141 CGTGGTCTCAGCTGAGAGCGCGCCGAGCAGAGAGAGAAAGAGGAGCGCCGAGAGCAG 1200
Db 1399 CGTGGTCTCAGCTGAGAGCGCGCCGAGCAGAGAGAGAAAGAGGAGCGCCGAGAGCAG 1458
QY 1201 GAGGCAAGCGGAGCTGAGAGCTGAGAAAGCAGCTGAGAAAGCAGCGGAGCTGAGAGC 1260
Db 1459 GAGGCAAGCGGAGCTGAGAGCTGAGAAAGCAGCTGAGAAAGCAGCGGAGCTGAGAGC 1518
QY 1261 CAGCGAGAGAGAGAGAGAGAGAGATCGAAGGCGCGAGCGCGCAAAACGGAAATC 1320
Db 1519 CAGCGAGAGAGAGAGAGAGAGAGATCGAAGGCGCGAGCGCGCAAAACGGAAATC 1578
QY 1321 GAAAGGCAAGCACTTTGAATGGGAAACGGAAACGGAAACCTCTGAAATCGAGAG 1380
Db 1579 GAAAGGCAAGCACTTTGAATGGGAAACGGAAACGGAAACCTCTGAAATCGAGAG 1638
QY 1381 AACCAAGAGCAGGAGGAGCAGCGGTCTCTGAAGGCAAGAGAGAACTCTGAGATTGAG 1440
Db 1639 AACCAAGAGCAGGAGGAGCAGCGGTCTCTGAAGGCAAGAGAGAACTCTGAGATTGAG 1698
QY 1441 TTGAAGCTCTGAATGACAAAAGATCAGCTAGAGGAAAACTTCAGATATCAGGTGT 1500
Db 1699 TTGAAGCTCTGAATGACAAAAGATCAGCTAGAGGAAAACTTCAGATATCAGGTGT 1758
QY 1501 CGACTGGCAACCAAGAGGCAAGAAATTTGAGAGCAGAACTCTGAGAGCTAAGAAAT 1560
Db 1759 CGACTGGCAACCAAGAGGCAAGAAATTTGAGAGCAGAACTCTGAGAGCTAAGAAAT 1818
QY 1561 GCTGAATCAACCACTTAACGAGCAGGTGCAAGAACTTCAGAAATGCTGGAGACTT 1620
Db 1819 GCTGAATCAACCACTTAACGAGCAGGTGCAAGAACTTCAGAAATGCTGGAGACTT 1878
QY 1621 ATTCCAGAGAAACAGATCTCAGTGAACAAGTTAAACAACTCCAGCAGAACTTTGCA 1680
Db 1879 ATTCCAGAGAAACAGATCTCAGTGAACAAGTTAAACAACTCCAGCAGAACTTTGCA 1938
QY 1681 AAGAGATCTGCTTTTACCTCAAAAGAGCTTTGAAAGCAAGAGCTGGCCGCGAGCAG 1740
Db 1939 AAGAGATCTGCTTTTACCTCAAAAGAGCTTTGAAAGCAAGAGCTGGCCGCGAGCAG 1998
QY 1741 CTCGAGAGAGCTGAGAGAGTGAAGAGAGAGAGAGAGAGCTCAAGCTGAGAGAGATTGAT 1800
Db 1999 CTCGAGAGAGCTGAGAGAGTGAAGAGAGAGAGAGAGAGCTCAAGCTGAGAGATTGAT 2058
QY 1801 GTTTTCAACAAACAGCTGAAGAACTGAGAGAGATACATAGCAAAACAGCACTCCAGAG 1860
Db 2059 GTTTTCAACAAACAGCTGAAGAACTGAGAGAGATACATAGCAAAACAGCACTCCAGAG 2118

QY 1861 CAGAGGTCCCTGAGAGCAGCGGACTGAAGAGCAAGAAAGACAGAGAGAGAGCGCTGAG 1920
Db 2119 CAGAGGTCCCTGAGAGCAGCGGACTGAAGAGCAAGAAAGACAGAGAGAGAGCGCTGAG 2178
QY 1921 TTGAAGAAACAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1980
Db 2179 TTGAAGAAACAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2238
QY 1981 GAGCATGTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2040
Db 2239 GAGCATGTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2298
QY 2041 AAGAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2100
Db 2299 AAGAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2358
QY 2101 GACAAAGCAGAGTCCGCTTTTCCATCCGATCAGAGAGAGAGAGAGAGAGAGAGAGAG 2160
Db 2359 GACAAAGCAGAGTCCGCTTTTCCATCCGATCAGAGAGAGAGAGAGAGAGAGAGAGAG 2418
QY 2161 CCTGGTCTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2220
Db 2419 CCTGGTCTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2478
QY 2221 GATATATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2280
Db 2479 GATATATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2538
QY 2281 CCAGAGAGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2340
Db 2539 CCAGAGAGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2598
QY 2341 GAGCTGAAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2400
Db 2599 GAGCTGAAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2658
QY 2401 GAGGTTCCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2460
Db 2659 GAGGTTCCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2718
QY 2461 GCTTGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2520
Db 2719 GCTTGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2778
QY 2521 AACCACTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2580
Db 2779 AACCACTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2838
QY 2581 GACCACTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2640
Db 2839 GACCACTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2897
QY 2641 CGGAGAGATCAGAGCTTTAACCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2700
Db 2899 CGGAGAGATCAGAGCTTTAACCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2957
QY 2701 GGCAGAGGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2760
Db 2959 GGCAGAGGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3017
QY 2761 AAAAGCAACCACTTTAAATTTTAAACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2820
Db 3018 AAAAGCAACCACTTTAAATTTTAAACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3077
QY 2821 ATGTGTGTGTTGAGAGAGTTCAAGGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2880
Db 3078 ATGTGTGTGTTGAGAGAGTTCAAGGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3137
QY 2881 CTCATTTCAAGGAGCGCTGAAGAGAAATCAACAGATGATCTGAGCTTAAGTAAAGTCT 2940
Db 3138 CTCATTTCAAGGAGCGCTGAAGAGAAATCAACAGATGATCTGAGCTTAAGTAAAGTCT 3197

```

QY 2241 GCTAGCTAAAGAGAGTGGCTTCCCGGCGCCGCAAGCAGCATTTCCGGAGAAAGATT 3000
    |||
Db 3198 GCTAGCTAAAGAGAGTGGCTTCCCGGCGCCGCAAGCAGCATTTCCGGAGAAAGATT 3257
    |||
QY 3001 ATTGCCATGTACACATACGAGAGTTCTGAGCAAGAGATTAACTTTACCTTTACAGCAAGGGAT 3060
    |||
Db 3258 ATTGCCATGTACACATACGAGAGTTCTGAGCAAGAGATTAACTTTACAGCAAGGGAT 3317
    |||
QY 3061 GTGATTTGGTTTACCAAGAAAGATGTGATCTGGTGGAGGGGAAAGGGGGGCAAGTCC 3120
    |||
Db 3318 GTGATTTGGTTTACCAAGAAAGATGTGATCTGGTGGAGGGGAAAGGGGGGCAAGTCC 3377
    |||
QY 3121 GGAAGTCTTCCCTTCTAATATGTGAGGCTTAAAGTTTCAAGAGGGCTCTGAAATGCTGGG 3180
    |||
Db 3378 GGAAGTCTTCCCTTCTAATATGTGAGGCTTAAAGTTTCAAGAGGGCTCTGAAATGCTGGG 3437
    |||
QY 3181 AAAACAGGAGTTTNGAAAAAAACCTGAATTTGCCAGTTATTGCTTCTTCTAGGCTGCT 3240
    |||
Db 3438 AAAACAGGAGTTTNGAAAAAAACCTGAATTTGCCAGTTATTGCTTCTTCTAGGCTGCT 3497
    |||
QY 3241 ACTGCTCCGCAACATCACTGCTGCTGCGGAGCTGATTTCTGATCCGGAAAAAGAAC 3300
    |||
Db 3498 ACTGCTCCGCAACATCACTGCTGCTGCGGAGCTGATTTCTGATCCGGAAAAAGAAC 3557
    |||
QY 3301 CCAGGTGATGATGTTGGGAAGAGAACTGCAAGCTGAGGAGAAAAAGCCAGATAGGGTGG 3360
    |||
Db 3558 CCAGGTGATGATGTTGGGAAGAGAACTGCAAGCTGAGGAGAAAAAGCCAGATAGGGTGG 3617
    |||
QY 3361 TTTCCAGCAAAATTATGTCAAACTTTTAAAGCCCGGAAACAGCAAAATACCCCAACTGAG 3420
    |||
Db 3618 TTTCCAGCAAAATTATGTCAAACTTTTAAAGCCCGGAAACAGCAAAATACCCCAACTGAG 3677
    |||
QY 3421 CTACCCAGAACCGGAGTGCAGCAGCAGTGTGCGCAGGTGATCGGGATGTAGATTAACAC 3480
    |||
Db 3678 CTACCCAGAACCGGAGTGCAGCAGCAGTGTGCGCAGGTGATGTAGATTAACAC 3737
    |||
QY 3481 GCCCAGAAAGATGAGCAACTAGCTTTCAGCAAAAGGCCAGATCATCAAGCTCTCAACAG 3540
    |||
Db 3738 GCCCAGAAAGATGAGCAACTAGCTTTCAGCAAAAGGCCAGATCATCAAGCTCTCAACAG 3797
    |||
QY 3541 GAGGACCCGAGCTGTGAGAAAGAGAGAGTCAAGTGGGGCAAGTTGGCTCTTCCCATCAAT 3600
    |||
Db 3798 GAGGACCCGAGCTGTGAGAAAGAGAGAGTCAAGTGGGGCAAGTTGGCTCTTCCCATCAAT 3857
    |||
QY 3601 TATGTAAAGCTGACCAAGACATGAGACCCCAAGCCAGCAATGA 3642
    |||
Db 3858 TATGTAAAGCTGACCAAGACATGAGACCCCAAGCCAGCAATGA 3899
    |||

```

RESULT 5
AAZ34572
ID AAZ34572 standard; cDNA; 5195 BP.

AAZ34572;

DT 01-FEB-2000 (first entry)

DE Human SH3D1A cDNA clone 11.

SH3D1A gene; human; Down's syndrome; leukaemia; cancer;
megakaryocytic abnormality; myeloproliferative disorder;
platelet disorder; neural disorder; thrombocytopenia;
haematopoietic disorder; cognitive dysfunction; microcephaly;
liisencephaly; colpocephaly; holoprosencephaly; diagnosis; therapy;
88.

OS Homo sapiens.

Key Location/Qualifiers

FT CDS 239..3886 /*tag= a

PN MO9953062-A2.

```

XX 21-OCT-1999.
PD 16-APR-1999; 99NO-US08371.
XX 16-APR-1999; 99NO-US08371.
PR 16-APR-1998; 98US-0082007.
XX (CEDA-) CEDARS SINAI HEALTH SYSTEM.
PA Korenberg JR, Chen X;
PI WPI; 1999-633829/54.
DR P-PSDB; AAY32156.
XX Nucleic acid from the human SH3D1A gene and its products, useful for
PT the diagnosis and treatment of myeloproliferative disorders and
PT leukaemia -
PS Claim 2; Fig 10; 99p; English.
XX This is the nucleotide sequence of full-length cDNA (clone 11)
CC corresponding to a novel human SH3 gene, termed the SH3D1A gene,
CC that contributes to the development of platelets and the
CC pathogenesis of leukaemias, both in general and in particular those
CC involving the megakaryocytic lineage. The SH3D1A gene maps to the
CC small candidate region for low platelets on chromosome 21.
CC Sequencing of 5 different sizes of cDNA clone from foetal brain
CC (see AAZ34570-74) suggests that at least 3 isoforms exist. The
CC invention provides methods for the diagnosis and treatment of
CC megakaryocytic abnormality, myeloproliferative disorder, platelet
CC disorder, acute leukaemia, neural disorders, thrombocytopenia,
CC platelet disorder on chromosome 21, low platelets in deletion for
CC 21, association of gains in chromosome 21 with leukaemias, neural
CC abnormalities, dysfunctions and disorders including brain
CC malformations and corresponding cognitive dysfunctions, are also
CC microcephaly, liisencephaly, and colpocephaly. Methods are also
CC provided for: suppressing cells unable to regulate themselves;
CC screening for a somatic alteration in the SH3D1A gene; monitoring
CC the progress and adequacy of a treatment; monitoring tumour risk
CC progress or megakaryocytic abnormality, myeloproliferative disorder,
CC haematopoietic disorder, platelet disorder or leukaemia, and
CC treatment of a subject (including a prenatal subject) having
CC megakaryocytic abnormality, myeloproliferative disorder,
CC platelet disorder, leukaemia or neural disorder using a
CC nucleic acid that expresses SH3D1A or its antisense nucleic acid.
XX Sequence 5195 BP; 1592 A; 1148 C; 1237 G; 1218 T; 0 other;
SQ
Query Match 78.4%; Score 2856; DB 20; Length 5195;
Best Local Similarity 86.8%; Pred. No. 0;
Matches 3167; Conservative 0; Mismatches 475; Indels 6; Gaps 2;
QY 1 ATGGCTCAAGTTTCCCAACCTTTGGTGTAGCTGATGTCTGGCCATACCTGTGAG 60
    |||
Db 239 ATGGCTCAAGTTTCCCAACCTTTGGTGTAGCTGATGTCTGGCCATACCTGTGAG 298
    |||
QY 61 GAAAGGGCGCAAGCATGACCAAGTCCCTTAAAGCCGATGAGGGGATTATTACT 120
    |||
Db 299 GAAAGGGCGCAAGCATGACCAAGTCCCTTAAAGCCGATGAGGGGATTATTACT 358
    |||
QY 121 GGTATCAAGCGAAGAACTTTTTCATCTGGGTACTCAAGCTGTCTTAGCAAA 180
    |||
Db 359 GGTATCAAGCGAAGAACTTTTTCATCTGGGTACTCAAGCTGTCTTAGCAAG 418
    |||
QY 181 ATATGGGCGCTAGCGGACATGATACGATGAGAAATGATCATCAAGTGAATTTTCCATA 240
    |||
Db 419 ATATGGGCGCTAGCGGACATGATGATGAGAAATGATGATCAAGTGAATTTTCCATA 478
    |||
QY 241 GCCATGAAGCTTATCAATCAAGTCAAGATATACGCTCCCTCCACACTCCCTCCCT 300
    |||
Db 479 GCTATGAAGCTTATCAATCAAGTCAAGATATACGCTCCCTCCACACTCCCTCCCT 538
    |||
QY 301 GTCATGAAGCAAGCAAGTGGCTATTTCAGTGCACAGCAATTTGGTATGAGAGGATT 360
    |||

```

Db 539 GTTCATGAACAGAACCCAGTTGCTATTTCTAGCGCACAGCATTTGTGTGGAGGTATC 598
 Qy 361 GGTAGATGCCACCACTCAACAGTGTGCTCTGTGTCCCAATGGGCTCCATTCAGTTGTT 420
 Db 599 GCCAGATGCCACCGCTTACAGCTGTGTGCTCCAGTCCAAATGGGAATCAATTCAGTTGTT 658
 Qy 421 GGAATGTCTCCACCCCTTAGTATCTTCTGTCCCTCCAGCAGCAGTGTGCTCCCTGGCTAAC 480
 Db 659 GGAATGTCTCCACCCCTTAGTATCTTCTGTCCCAAGCAGCAGTGTGCTCCCTGGCTAAC 718
 Qy 481 GGAGGTCTCTCCCTGTATACAGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT 540
 Db 719 GGAGGTCTCTCCCTGTATACAGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT 778
 Qy 541 AAGAGTTCTCTCTGTATACAGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT 600
 Db 779 AAGAGTTCTCTCTGTATACAGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT 838
 Qy 601 GCAATCATTCGATGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT 660
 Db 839 GCAATCATTCGATGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT 898
 Qy 661 TCAAGGCTAAATACAGGAGATTATTCACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 720
 Db 899 TCAAGGCTAAATACAGGAGATTATTCACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 958
 Qy 721 AAGAGTCTCTCTGTATACAGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT 780
 Db 959 AAGAGTCTCTCTGTATACAGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT 1018
 Qy 781 TCAATGTGATCTTCTGTATACAGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT 840
 Db 1019 TCAATGTGATCTTCTGTATACAGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT 1078
 Qy 841 CTAGCTATGCTCTCTGTATACAGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT 900
 Db 1079 CTAGCTATGCTCTCTGTATACAGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT 1138
 Qy 901 CCAGATATATCTCTCTGTATACAGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT 960
 Db 1139 CCAGATATATCTCTCTGTATACAGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT 1198
 Qy 961 AGCTCTCTCTGTGTGATGAGGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT 1020
 Db 1199 AGCTCTCTCTGTGTGATGAGGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT 1258
 Qy 1021 C---CAGAGAGAACTGCTGTGATTCATTTGAATAGAGAGGAGAGAGAGAGAGAGAGAGAG 1077
 Db 1259 CAATTAGAAAGAAATTTACCTGTAACTTTGAAGATTAAGAGGAGAGAGAGAGAGAGAGAG 1318
 Qy 1078 GGCAGTGTGAGCTGTGAGAGAGGCGCCCAAGCGCTTTTGAGAGCAGAGCGCAAGAGCAG 1137
 Db 1319 GGCAGTGTGAGCTGTGAGAGAGGCGCCCAAGCGCTTTTGAGAGCAGAGCGCAAGAGCAG 1378
 Qy 1138 GAGCGTGTGCTGTGAGAGGCGCCCAAGCGCTTTTGAGAGCAGAGCGCAAGAGCAGAG 1197
 Db 1379 GAGCGTGTGCTGTGAGAGGCGCCCAAGCGCTTTTGAGAGCAGAGCGCAAGAGCAGAG 1438
 Qy 1198 CAGAGGCGCAAGCGCTGTGAGAGAGCAGCTGTGAGAGCAGAGCAGAGCAGAGCAGAG 1257
 Db 1439 CAGAGGCGCAAGCGCTGTGAGAGAGCAGCTGTGAGAGCAGAGCAGAGCAGAGCAGAG 1498
 Qy 1258 CCGCAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1317
 Db 1499 CCGCAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1558
 Qy 1318 CTGGAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1377
 Db 1559 CTGGAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1618
 Qy 1378 AGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1437

Db 1619 AGAAACAAAGACAAAGAGACATAGTTGTACTGAAAGCAAAAGAAAAAGACTTTGGAATTT 1678
 Qy 1438 GAGTTAGAGCTCTGAAATGACAAAGAAAGCTAGCTTAGAAGAAAGAACTTAGATATCAG 1497
 Db 1679 GAAATAGAGCTCTAAATGATAAAAAGATCACTAGAAAGGAAAGAACTTAGATATCAG 1738
 Qy 1498 TGTGACTGGCAAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1557
 Db 1739 TGTGACTGACCAAGCCAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1798
 Qy 1558 ATTGCTGAAATACCCACTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1617
 Db 1799 ATTGCTGAAATACCCACTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1858
 Qy 1618 CTTATTTCCAGAGAAACAGATATCTAGTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1677
 Db 1859 CTTATTTCCAGAGAAACAGATATCTAGTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1918
 Qy 1678 CATAGAGACTGCTCTTACCTCAAAAGAGCTTGAAGCAAAAGAGAGAGAGAGAGAGAGAGAG 1737
 Db 1919 CATAGAGATTCCTTGTACTTAAAGAGCTTGAAGCAAAAGAGAGAGAGAGAGAGAGAGAGAG 1978
 Qy 1738 CAGCTCCGAGAGCAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1977
 Db 1979 CAGCTACAGAGCAAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2038
 Qy 1798 GATGTTTTCAACACAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1857
 Db 2039 GATATTTTCATTAATCAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2098
 Qy 1858 AAGCAGAGAGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1917
 Db 2099 AAGCAGAGAGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2158
 Qy 1918 GAGTTAGAGAGCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1977
 Db 2159 GAGTTAGAGAGCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2218
 Qy 1978 CTGAGAGATGTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2034
 Db 2219 CTGAGAGATGTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2278
 Qy 2035 AGACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2094
 Db 2279 AGACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2338
 Qy 2095 ATGCAAGACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2154
 Db 2339 GCACAAAGACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2398
 Qy 2155 CAGGCACTGTGTTTACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2214
 Db 2399 CAGGCACTGTGTTTACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2458
 Qy 2215 AAGAGTATATTTACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2274
 Db 2459 AAGAGTATATTTACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2518
 Qy 2275 ATCCAGCAGAGAGATATAGTCAATGTGTGATGAAAGCAGAGAGAGAGAGAGAGAGAGAG 2334
 Db 2519 ATCCAGCAGAGAGATATAGTCAATGTGTGATGAAAGCAGAGAGAGAGAGAGAGAGAGAG 2578
 Qy 2335 GAGAGAGAGCTGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2394
 Db 2579 GAGAGAGAGATTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2638
 Qy 2395 GAAATAGAGGTTCCACTCCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2454
 Db 2639 GAAATAGAGGTTCCACTCCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2698
 Qy 2455 AACTGAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2514
 Db 2699 AACTGAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2758

```

OY 2515 ACCCCCAACAATGCGGAGACTTCAGTTCACAGTGGCCAGAGCTCAAGAGAGCA 2574
DB 2759 ACCCTTAATTAATCTGGCCGAGCTTCAGCTCAGTGGCCACAGACAGAAAGAGAAACA 2818
OY 2575 GAAACGAGCAACTGGGATACGTGGCGGCTCAGCCCTTCTGACCGTACTAGTGTGGC 2634
DB 2819 GAAACGATTAATCTGGGATGATAGGCGACGCCCTCTCTACCGTTCCAAAGTGGCGGC 2878
OY 2635 CAGTTACGGCAGAGATCAGCCTTTACCCAGCCACAGCCACTGGTCTCTCCCATCTCC 2694
DB 2879 CAGTTAAGCAGAGGTCCGCTTTACTCAGCCAGCGCCACTGGCTCTCCCGCTCTCT 2938
OY 2695 GTCCGGGCGCAGGGGAGAAAAGGTGAGAGGGCTCAAGCGCAAGCCCTGATCCCGGAGA 2754
DB 2939 GTGCTAGGCGCAGGGTGAAGAGGTGAGAGGGCTCAAGCTCAAGCCCTTATCTTGGAGA 2998
OY 2755 GCCAAAAAAGACACACACTTAATTTTAAACAAGTGAAGTCAACCGTCTGAGACAG 2814
DB 2999 GCCAAAAAAGACACACTTAATTTTAAACAAGTGAAGTCAACCGTCTGAGACAG 3058
OY 2815 CAAGCATGTGTGTTGGAGAGTTCAAGTCAAGAGGGTGGTCCCAAGCTTAC 2874
DB 3059 CAAGCATGTGTGTTGGAGAGTTCAAGTCAAGAGGGTGGTCCCAAGCTTAC 3118
OY 2875 GTGAACATCTTTGAGGCGCCGTAGAGAAATCAACAAGCTGATACCTGGCCCTAGAA 2934
DB 3119 GTGAACATCTTTGAGGCGCCGTAGAGAAATCAACAAGCTGATACCTGGCTTCAAGG 3178
OY 2935 AGTCTGTAGTCTTAAGAGAGTGTCTCCCGCGCGCAAGCCACTTCCCGAGAA 2994
DB 3179 AGTCTGTAGTCTTAAGAGAGTGTCTCTCCAGAGCAAGCCGCTTGGGAGAA 3238
OY 2995 GAGTTATTCAGTATCAATACAGAGGTTTGAGCAAGAGATTTAACTTTAGACA 3054
DB 3239 GAAATTAATTCAGTATCAATACAGAGGTTTGAGCAAGAGATTTAACTTTAGACA 3298
OY 3055 GGGGATGTGATTTGTTTACCAAGAAAGTGTGCTGGTGAAGGGAACGGTGGGCGAC 3114
DB 3299 GGGGATGTGATTTGTTTACCAAGAAAGTGTGCTGGTGAAGGGAACGGTGGGCGAC 3358
OY 3115 AAGTCGGAGTCTTCCCTTCTAATATGAGAGGCTTAAAGATTCAAGAGGCTCTGAACT 3174
DB 3359 AAGTCGGAGTCTTCCCTTCTAATATGAGAGGCTTAAAGATTCAAGAGGCTCTGAACT 3418
OY 3175 GCTGGGAAAACAGGAGTTTAGAAAAAACTGAAATTTGCCAGGTTATGCTTCTAC 3234
DB 3419 GCTGGGAAAACAGGAGTTTAGAAAAAACTGAAATTTGCCAGGTTATGCTTCTAC 3478
OY 3235 GCTGTACTGTGTCGGGAAACAATCAACCTGGCTCTGGGCGAGTGAATCTGATCCGGAA 3294
DB 3479 ACCGCAACCGGCGCCGAGAGCTCACTCTCGCCCTGTGCTGATATTTGATCCGAAA 3538
OY 3295 AAGAACCCAGGTGTGATGTGGAGAGAACTGCAAGCTCGAGGAAAAAGCGCCAGATA 3354
DB 3539 AAGAACCCAGGTGTGATGTGGAGAGAGCTGCAAGCACTGGGAAAAAGCGCCAGATA 3598
OY 3355 GGGTGTGTTTCCAGCAATTAATATCTCAAACTTTAAGCCCGGAAACAAGAAATCAACCA 3414
DB 3599 GGGTGTGTTTCCAGCAATTAATATCTCAAACTTTAAGCCCGGAAACAAGAAATCAACCA 3658
OY 3415 ACTGAGCTAACCCAAACCGCAGTGAAGCAGAGAGTGTGCAAGTATGGGGATGACAGT 3474
DB 3659 AAGAGGCCAATTAATCAACACATTAAGGAGAGTGTGCAAGTATGGGGATGACAGT 3718
OY 3475 TACACCGCCCAAGAACGATGACGATAGCTTCAAGCAAGGCGCAGATATCAACGTCCTC 3534
DB 3719 TACACCGCGCAGAAATGACGATAGCTGCTTCAACAAAGGCGCAGATATCAACGTCCTC 3778
OY 3535 AACAAAGAGAGACCCGAGCTGTGTGAAAGAGAAAGTCAAGTGGCAAGTTGGGCTTTCCCA 3594
DB 3779 AACAAAGAGAGACCCGAGCTGTGTGAAAGAGAAAGTCAATGCAAGTGGGCGCTTTCCCA 3838

```

```

OY 3595 TCCAATTATGTAAAGCTGACACAGACATGAGACCCAGCCAGCAATGA 3642-
DB 3839 TCCAATTATGTGAAGCTGACACAGACATGAGACCCAGCCAGCAATGA 3886

RESULT 6
ID AA234571 standard; cDNA, 5458 BP.
XX AA234571;
AC 01-FEB-2000 (first entry)
XX
DE Human SH3D1A cDNA clone 21.
XX
KW SH3D1A gene; human; Down's syndrome; leukaemia; cancer;
KW megakaryocytic abnormality; myeloproliferative disorder;
KW platelet disorder; neural disorder; thrombocytopenia;
KW haematopoietic disorder; cognitive dysfunction; microcephaly;
KW lissencephaly; colpocephaly; holoprosencephaly; diagnosis; therapy;
KW s8.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 267..3929
FT /tag= a
FT
XX
EN MO9953062-A2.
XX
PD 21-OCT-1999.
XX
PE 16-APR-1999; 99WO-US08371.
XX
PR 16-APR-1998; 98US-0082007.
XX
PA (CEDA-) CEDARS SINAI HEALTH SYSTEM.
XX
PI Korenberg JR, Chen X;
XX
DR WP1; 1999-633829/54.
XX
P-PSDB; AAY32155.
XX
PT Nucleic acid from the human SH3D1A gene and its products, useful for
PT the diagnosis and treatment of myeloproliferative disorders and
PT leukaemia.
XX
PS Claim 2; Fig 8; 99p; English.
XX
CC This is the nucleotide sequence of full-length cDNA (clone 21)
CC corresponding to a novel human SH3 gene, termed the SH3D1A gene,
CC that contributes to the development of platelets and the
CC pathogenesis of leukaemias, both in general and in particular those
CC involving the megakaryocytic lineage. The SH3D1A gene maps to the
CC small candidate region for low platelets on chromosome 21.
CC Sequencing of 5 different sizes of cDNA clone from foetal brain
CC (see AA234570-74) suggests that at least 3 isoforms exist. The
CC invention provides methods for the diagnosis and treatment of
CC megakaryocytic abnormality, myeloproliferative disorder, platelet
CC disorder, acute leukaemia, neural disorders, thrombocytopenia,
CC platelet disorder on chromosome 21, low platelets in deletion for
CC 21, association of gains in chromosome 21 with leukaemias, neural
CC abnormalities, dysfunctions and disorders including brain
CC malformations and corresponding cognitive dysfunctions,
CC microcephaly, lissencephaly, and colpocephaly. Methods are also
CC provided for: suppressing cells unable to regulate themselves;
CC screening for a somatic alteration in the SH3D1A gene; monitoring
CC the progress and adequacy of a treatment; monitoring tumour risk
CC progress or megakaryocytic abnormality, myeloproliferative disorder,
CC haematopoietic disorder, platelet disorder or leukaemia; and
CC treatment of a subject (including a prenatal subject) having
CC megakaryocytic abnormality, myeloproliferative disorder,
CC platelet disorder, leukaemia or neural disorder using a

```



```

Db 2307 AAACGAAAGGAGAGAGTGTCAAAAAGAAAGATGGGAGAGAAAAGGCAACAGAA 2366
Qy 2095 ATGCAAGCAAGCAAGATCGGCTTTTCCATCCGATCGAGAGCCAGCTTAAGTGGCCAC 2154
Db 2367 GCACAAAGCAAGCTGGTGGCTTTTCCATCAACCAAGAAACGAGTAAAGCAGCTTTC 2426
Qy 2155 CAGGACCTGGTGTACCAAGAAAGAGCCGCTTACCATTTCTGACAGAGAGTGA 2214
Db 2427 CAGGACCTGGTGTACCAAGAAAGAGTTCATTCATTTCTGACAGAGAAATTA 2486
Qy 2215 AAAGGTATATTAACGAGCGCTGTACCCCTTGAATCGAAGATCAGATGATCAC 2274
Db 2487 AAAGGTATATTAACGAGCGCTGTACCCCTTGAATCGAAGATGATGATGATCACT 2546
Qy 2275 ATCCAGCCAGAGATATAGTCACT-----GTTGATGAAAGCCAGCTGGA 2319
Db 2547 ATCCAGCCAGAGATATAGTCACTGTTAAAGGGAATGGGTGATGAAAGCCAACTGGA 2606
Qy 2320 GAGCCAGATGGCTGAGAGAGCTGAAAGGAGAGAGGATGGTTCCTGCAAACTAT 2379
Db 2607 GAGCCAGCTGGCTTGAAGAGAAATTAAGAAAGAGAGAGGATGGTTCCTGCAAACTAT 2666
Qy 2380 GCAGAAAGATTCAGAAAGATGAGTTCCTCACTCAGCCAAACAGTGAACGATCTGA 2439
Db 2667 GCAGAAAGATTCAGAAAGATGAGTTCCTCACTCAGTGAACGATCTGATTCACA 2726
Qy 2440 TCTGCCCCCTGCCCCAACTGGCTTGGGTGAGACCCCTGCTTCTTTCAGTGA 2499
Db 2727 TCTGCCCCCTGCCCCAACTGGCTTGGGTGAGACCCCTGCTTCTTTCAGTGA 2786
Qy 2500 TCTGAGCCCTCCAGAACCCCAACTGGGAGAGCTTCACTTCAAGTGGGCGGAGAG 2559
Db 2787 TCAAGGCTCTCAGACACCTCTTAATCTGGGCGGAGCTTCAAGTGGGCGGAGAG 2846
Qy 2560 TCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2619
Db 2847 AGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2906
Qy 2620 GTACCTAGTGTGCGGAGATACGAGAGATACGCTTTACCCAGCCAGAGAGAGAG 2679
Db 2907 GTTCCAAATGCGGAGAGATTAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2966
Qy 2680 TCTCCCTCCATCTCCGCTCTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2739
Db 2967 TCTCCCTCCATCTCCGCTCTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3026
Qy 2740 CTGATATCCCTGGAGAGCCAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2799
Db 3027 CTATATCTCTGGAGAGCCAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3086
Qy 2800 ACCGTTCTGGAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2859
Db 3087 ACCGTTCTGGAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3146
Qy 2860 TTCCCAAGTCTTAAGTAACTCAATTCAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 2919
Db 3147 TTCCCAAGTCTTAAGTAACTCAATTCAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 3206
Qy 2920 ACTGGCCCTACTGAAGTCTGCTAGTCTAAAGAGAGAGAGAGAGAGAGAGAGAGAG 2979
Db 3207 TCTGGTCTTGAAGAGTCTGCTAGTCTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3266
Qy 2980 GCCATTCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3039
Db 3267 GTCGTTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3326
Qy 3040 TTAACCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3099
Db 3327 TTAACCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3386
Qy 3100 GGAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3159
Db 3387 GGAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3446

```

```

Qy 3160 GAGGCTCTGGAACTGCTGGGAAAAACAGGAGTTTAGAAAAAAACCTGAAATTGGCCAG 3219
Db 3447 GAGGCTCTGGAACTGCTGGGAAAAACAGGAGTTTAGAAAAAAACCTGAAATTGGCCAG 3506
Qy 3220 GTTATGCTTCTTCAAGCTGCTACTGTGTCCGAAACACTCACTCTGGCTCTGGGAGCTG 3279
Db 3507 GTTATGCTTCTTCAAGCTGCTACTGTGTCCGAAACACTCACTCTGGCTCTGGGAGCTG 3566
Qy 3280 ATTCTGATCCGAAAAAGAACCCAGGTGATGTGGAGAGAGAACTGCAAGCTCGAGGG 3339
Db 3567 ATTCTGATCCGAAAAAGAACCCAGGTGATGTGGAGAGAGAACTGCAAGCTCGAGGG 3626
Qy 3340 AAAAGGCGCAGATAGAGGTGGTTTCCAGCAAAATTAATGCTTCAAGCCCGGAGACA 3399
Db 3627 AAAAGGCGCAGATAGAGGTGGTTTCCAGCTAATTAATGCTTCAAGCCCTGGGAGAG 3686
Qy 3400 AGCAAAATCAACCCCACTAGCTATCCCAAGAGCCGAGTGAAGCCAGAGTGTGCAAGT 3459
Db 3687 AGCAAAATCACTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3746
Qy 3460 ATCGGATGTACGATTAACAGCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3519
Db 3747 ATCGGATGTACGATTAACAGCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3806
Qy 3520 ATCAATCAAGCTCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3579
Db 3807 ATCAATCAAGCTCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3866
Qy 3580 GTTGGGCTCTTCCATCAATTAATGATGAAGCTGACAGAGAGAGAGAGAGAGAGAGAG 3639
Db 3867 GTTGGGCTCTTCCATCAATTAATGATGAAGCTGACAGAGAGAGAGAGAGAGAGAGAG 3926
Qy 3640 TGA 3642
Db 3927 TGA 3929

RESULT 7
AAS84763
ID AAS84763 standard; cDNA; 7435 BP.
XX AAS84763;
AC
XX
AC
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #20567.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
OS Homo sapiens.
XX
PN MO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
XX
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI: 2001-639362/73.
XX
DR P-FSDB: ABG20576.
XX
PT New isolated polynucleotide and encoded polypeptides; useful in
PT diagnostics, forensics; gene mapping; identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -

```


Claim 1; SEQ ID No 20567; 103bp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. AAS64197-AAS94564 represent novel human CC diagnostic coding sequences of the invention. Note: The sequence data for this patent did not appear in the printed CC specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pot_sequences.

Sequence 7435 BP; 2190 A; 1690 C; 1829 G; 1726 T; 0 other;

Query Match 68.4%; Score 2491.6; DB 23; Length 7435; Best Local Similarity 84.2%; Pred. No. 0; Matches 3112; Conservative 0; Mismatches 529; Indels 53; Gaps 25;

1 ATGGCTCAGTTTCCACACCTTGGTGGTAGCTGGAGTCTGGGSCATACTGGAG 60
 263 ATGGCTCAGTTTCCACACCTTGGTGGTAGCTGGAGTCTGGGSCATACTGGAG 322
 61 GAAAGGCGCAAGCATGACAGCAGTCTTGGCTGAGCCGATAGCGGGATTATTACT 120
 323 GAAAGGCGCAAGCATGACAGCAGTCTTGGCTGAGCCGATAGCGGGATTATTACT 382
 121 GGTGATCAAGCGAGAACTTTTTCATCTGGGTTACCTGAGCTGTCTTACAGCAA 180
 383 GGTGATCAAGCGAGAACTTTTTCATCTGGGTTACCTGAGCTGTCTTACAGCAA 442
 181 ATATGGGCGCTAGCGGACATGATACGATGGAAGATGATCAATGGAATTTTCCATA 240
 443 ATATGGGCGCTAGCGGACATGATGATGGAAGATGATCAATGGAATTTTCCATA 502
 241 GCCATGAAGCTTATCAAACTGAAAGCTACAGGATATCAGTCCCTCCACACTTCCCT 300
 503 GCTATGAAGCTTATCAAACTGAAAGCTACAGGATATCAGTCCCTCCACACTTCCCT 562
 301 GTTCATGAAGCAAGCAACAGTGGCTATTTCAAGTCCAGCAGCAATTGGTATGAGAGGATT 360
 563 GTTCATGAAGCAAGCAACAGTGGCTATTTCAAGTCCAGCAGCAATTGGTATGAGAGGATT 622
 361 GTTAGAGTCCCACTCAAGCTGTGCTGCTGTCGCAATGGGCTCCACTT--CGAATTGT 419
 623 GCGAGATGCAACCGCTTAAAGCTGTGCTGCTGTCGCAATGGGCTCCACTTGT 682
 420 TGGAAATGTCACACCTTAGTATCTTCTGCTCCAGCAGAGAGTCCCTCCCTGGCTAA 479
 683 TGGAAATGTCACACCTTAGTATCTTCTGCTCCAGCAGAGAGTCCCTCCCTGGCTAA 742
 480 GGGGGCTCTCCCGTCATACAGCTCTGCTGCTGGCTTGGCATCTGACAGCAATGGCC 539
 743 CGGGGCTCTCCCGTCATACAGCTCTGCTGCTGGCTTGGCATCTGACAGCAATGGCC 802
 540 AAGAGTTCTTCTTCAAGAGATCTGGTCCAGGGTCAACAATTAAACATTAAGTAAAGAA 599
 803 AAGAGTTCTTCTTCAAGAGATCTGGTCCAGGGTCAACAATTAAACATTAAGTAAAGAA 862
 600 GGCACATCATTTGATGTGCGCAGCGCCCTCC--AGCAGCAGATGGCTGTGCTCACT 658

863 GGCACATCATTTGATGTGCGCAGCGCTCCACCAAGTGGCAGAGTGGGCTTCTCACT 922
 659 CATCAAGCTGAAATATCA---GGCAGTATTTAAACAGCCAGCAAAAC--TATGATGG 713
 923 CATCAAGCTGAAATATCAAGCAAAATATTCATATGATATCAAAACCTTATGATGG 982
 714 ACAGTT--AACAGTCCCGAGCAAGAACTATTTCTATGCAATCAAGTTTACCCAGGCT 771
 983 ACAGTTTAAAGAGTCCCGAGCAAGAACTATTTCTATGCAATCAAGTTTACCAAGGCT 1042
 772 CAGCTGGCTTCAATATGAAATCTTTCTGACATTTGATCAAGATGAAATCTGACAGAA 831
 1043 CAGCTGGCTTCAATATGAAATCTTTCTGACATTTGATCAAGATGAAATCTGACAGAA 1102
 832 GAATTTATCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 891
 1103 GAATTTATCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1162
 892 GTCTGGCTCCAGAAATACATCCCTCTCTTCAAGAGATGCTCCGCGAGTGGAGT 951
 1163 GTCTGGCTCCAGAAATACATCCCTCTCTTCAAGAGATGCTCCGCGAGTGGAGTATA 1222
 952 TCCGTCATAGCTCTTCTTCTGATGATGATGATGATGATGATGATGATGATGATGAT 1011
 1223 TCTGTCATAGCTCAACATCTGTATGATGATGATGATGATGATGATGATGATGATGAT 1282
 1012 GAGCAGCAGC---CAGAGAAATGCTGCTGATGATGATGATGATGATGATGATGATGAT 1068
 1283 GAGCAGCAGC---CAGAGAAATGCTGCTGATGATGATGATGATGATGATGATGATGAT 1342
 1069 TTCGAGCAGCAGCAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGG 1128
 1343 TTTGAAAGTGGCACTTGGAACTGGAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1402
 1129 AAGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1188
 1403 AAGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1462
 1189 CCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1248
 1463 CCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1522
 1249 GAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1308
 1523 GAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1582
 1309 AAGCGGAACT--GGAAGGCAAGCGAACTTGAATGGGAAACCGAGACAGGAACT 1367
 1583 AAGAGGCAAGCTATGAAAGCACTGACCTGGAAGTGGAGAGCATTCAGAGTCCGAACT 1642
 1368 CCGTGAATCAGAGAAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1427
 1643 ACTAATCAAGAGAAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1702
 1428 TCTGGAATTTGAGTGAAGAGCTCTGAATGACAAAGAGATCAGTGAAGAGAACTTCA 1487
 1703 TTTGGAATTTGAGTGAAGAGCTCTGAATGATGAAGAGATCAGTGAAGAGAACTTCA 1762
 1488 GATATTCAGAGTGTGCACTGCGCAACCCAGAGGCAAGAAATTTGAGAGCAGCAATGCTAG 1547
 1763 AGATATTCAGAGTGTGCACTGCGCAACCCAGAGGCAAGAAATTTGAGAGCAGCAATGCTAG 1822
 1548 AGAGCTAAGATTTGCTG--AAATCAACCACTTACAGCAGAGTT--GCAAGATCTCAGC 1603
 1823 AGAGTTGAAGATTTGCGGAAATCAACCCATCTACAGCACTAATTTTCAATGGAATCTCAGC 1882
 1604 AAA---TGCTTGAAGAGCTTATTCAGAGAAACAGATCTCAGTGAACAGTTAAAGAG 1660
 1883 AAGATCTTTGAGAGAACTTATTCAGAGAAACAGATCTCAGTGAACAGTTAAAGAG 1942
 1661 TCCAGCAGAGAGAGTGGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1719
 1943 TTCAGCAGAGAGAGTGGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2002

KW platelet disorder; neural disorder; thrombocytopenia;
 KW haematopoietic disorder; cognitive dysfunction; microcephaly;
 KW lissencephaly; colpocephaly; holoprosencephaly; diagnosis; therapy;
 ss.

OS Homo sapiens.

Key Location/Qualifiers
 FT CDS 208..3642
 FT /*tag= a

PN MO9953062-A2.

XX 21-OCT-1999.

PD 16-APR-1999; 99MO-US08371.

PR 16-APR-1998; 98US-0082007.

XX (CEDA-) CEDARS SINAI HEALTH SYSTEM.

XX Korenberg JR, Chen X;

XX MPI, 1999-633829/54.

XX P-PSDB; AAY32154.

PT Nucleic acid from the human SH3D1A gene and its products, useful for
 PT the diagnosis and treatment of myeloproliferative disorders and
 PT leukaemia.

PS Claim 2; Fig 5; 99p; English.

XX This is the nucleotide sequence of full-length cDNA corresponding
 CC to a novel human SH3 gene, termed the SH3D1A gene, that contributes
 CC to the development of platelets and the pathogenesis of leukemias,
 CC both in general and in particular those involving the
 CC megakaryocytic lineage. The SH3D1A gene maps to the small
 CC candidate region for low platelets on chromosome 21. Sequencing
 CC of 5 different sizes of cDNA clone (see AYZ34570-74) suggests that
 CC at least 3 isoforms exist. The invention provides methods for the
 CC diagnosis and treatment of megakaryocytic abnormality,
 CC myeloproliferative disorder, platelet disorder, acute leukaemia,
 CC neural disorders, thrombocytopenia, platelet disorder on
 CC chromosome 21; low platelets in deletion for 21, association of
 CC gains in chromosome 21 with leukemias, neural abnormalities,
 CC dysfunctions and disorders including brain malformations and
 CC corresponding cognitive dysfunctions, microcephaly, lissencephaly,
 CC and colpocephaly. Methods are also provided for: suppressing
 CC cells unable to regulate themselves; screening for a somatic
 CC alteration in the SH3D1A gene; monitoring the progress and
 CC adequacy of a treatment; monitoring tumour risk progress or
 CC megakaryocytic abnormality; myeloproliferative disorder,
 CC haematopoietic disorder, platelet disorder or leukaemia; and
 CC treatment of a subject (including a prenatal subject) having
 CC megakaryocytic abnormality, myeloproliferative disorder,
 CC platelet disorder, leukaemia or neural disorder using a
 CC nucleic acid that expresses SH3D1A or its antisense nucleic acid.

XX Sequence 5199 BP; 1615 A; 1152 C; 1214 G; 1218 T; 0 other;

Query Match 66.7%; Score 2429.6; DB 20; Length 5199;
 Best Local Similarity 81.1%; Pred. No. 0;
 Matches 2960; Conservative 0; Mismatches 469; Indels 219; Gaps 3;

QY 1 ATGGCTCAAGTTCCACACCTTGGTGTGAGCTGGATGTCTGGCCATTAACCTGTGAG 60
 DB ATGGCTCAAGTTCCACACCTTGGTGTGAGCTGGATGTCTGGCCATTAACCTGTGAG 267
 QY 61 GAAAGGGCAAGATACCAAGCAGTTCCTTAAAGCCATTAACCGGATTTATTACT 120
 DB GAAAGGGCAAGATACCAAGCAGTTCCTTAAAGCCATTAACCGGATTTATTACT 327
 QY 121 GGTGATCAAGCAGAACTTTTTCATCTGGGTTACCTCAGCTGTCTTAGACAA 180

DB 328 GGTGATCAAGCTAGAAATCTTTTTCATCTGGGTTACCTCAGCTGTCTTAGACAG 387
 QY 181 ATATGGGCGCTAGCGGACATGAATAAGATGAGATGATGAAATTTTTCATA 240
 DB ATATGGGCACTAGCTGACATGAATAATGAGAAAGATGATGAAATGATTTTTCATA 447
 QY 241 GCATGAGCTTATCAAACTGAAGGTACAAAGATATCACTCCCTCCACACTTCCCT 300
 DB GCTATGAACCTTATCAAACTGAAGGTACAAAGATATCACTCCCTCCACACTTCCCT 507
 QY 301 GTCATGAAACAGCAACCAAGTGCCTATTTCCAGTGCACAGCAATTTGTATGAGGAT 360
 DB GTCATGAAACAGCAACCAAGTGCCTATTTCCAGGACACAGCAATTTGTATGAGGAT 567
 QY 361 GCTAGATGCCACACTACAGCTGTGCTCTCTGTCATATGGGCTCCATTCAGTTGT 420
 DB GCAGCATGCCACCGCTTACAGCTGTGCTCCAGTGCATATGAGGATTCATTCAGTTGT 627
 QY 421 GGAATGTCCACACCTTAGTATCTTCTGTCCTCCAGCAGCAGTGCCTCCGCTAC 480
 DB GGAATGTCCACACCTTAGTATCTTCTGTCCTCCAGCAGCAGTGCCTCCGCTAC 687
 QY 481 GGGGCTCTCCGCTATACAGCTCTGCTGCTGCTTGGCCGATCTGCAACATGAGCA 540
 DB GGGGCTCCCGCTTATACAGCTCTGCTGCTGCTTGGCCGATCTGCAACATGAGCA 747
 QY 541 AAGATTTCTTCTTACAGCATGTGTCAGGCTCAATTAACATTAAGTTACAGAG 600
 DB AAGATTTCTTCTTATGATGTGTCAGGCTCAATTAACATTAAGTTACAGAG 807
 QY 601 GCACATCATTTGATGTGCGCAGCCCTCCAGCAGCAAGATGAGCTGTGCTGCTA 660
 DB GCACATCATTTGATGTGCGCAGCTCCAGCAGCAAGATGAGCTGTGCTGCTA 867
 QY 661 TCAAGGCTGAATATACAGCAGTATTAACAAGCAGCAAACTATGATGAGCACTTA 720
 DB TCAAGGCTGAATATACAGCAGTATTAACAAGCAGCAAACTATGATGAGCACTTA 927
 QY 721 AAGGTCCTCCAGCAGCAAGTATTCATGATCAATCAATCCAGGCTCAGCTGCT 780
 DB AAGGTCCTCCAGCAGCAAGTATTCATGATCAATCAATCCAGGCTCAGCTGCT 987
 QY 781 TCAATATGAAATCTTTCGACATGATCAAGATGAGAACTACGTCAGAAATATATC 840
 DB TCAATATGAAATCTTTCGACATGATCAAGATGAGAACTACGTCAGAAATATATC 1047
 QY 841 CTAGCTATGACCTATATGATGTGCAATGTCTGTCAGCAGCTCCGCTGCTGCT 900
 DB CTAGCTATGACCTATATGATGTGCAATGTCTGTCAGCAGCTCCGCTGCTGCT 1107
 QY 901 CCAAGATATACCTCTCTCTTCCAGAAAGTTCCTCCGAGTGGATGCTGCTAT 960
 DB CCAAGATATACCTCTCTCTTCCAGAAAGTTCCTCCGAGTGGATGCTGCTAT 1167
 QY 961 AACTCTTCTTCTGATATGAGGCTGCTGAGAGCCGTCTCAGAGATGAGCAGCAG 1020
 DB AACTCTTCTTCTGATATGAGGCTGCTGAGAGCCGTCTCAGAGATGAGCAGCAG 1227
 QY 1021 C---CAGAGAAAGAACTGCTGACATTTGAAGATTAAGAAAGCGGAGAACTTGA 1077
 DB CATTATGAAAGAAATTAACCTTTAAGATTTGAAGATTAAGAAAGCGGAGAACTTGA 1287
 QY 1078 GGCAGTGTGAGCTGAGAGAGCGCCGCAAGGCTCTTGTGAGCAGCAGGCAAGAGCAG 1137
 DB GGCAGTGTGAGCTGAGAGAGCGCCGCAAGGCTCTTGTGAGCAGCAGGCAAGAGCAG 1347
 QY 1138 GAGCGTTGCTTCTACCTGAGAGCGCCGCAAGCAGAGAGAAAGACCGGAGCGCAGAG 1197
 DB GAGCGTTGCTTCTACCTGAGAGCGCGGCAAGCAGAGAGAAAGACCGTATGAGCGCAGAG 1407
 QY 1198 CAGAGAGCGCAAGCGCAGCTGAGCTGAGAGAGCAGCTGAGAGAGCAGCGGAGCTGAG 1257

QY 3415 ACTGAGCTACCCAGACCCGAGTGCAGCCAGCAGTGTCCAGTATCCGAGTGTACGAT 3474
DB 3415 ACAGAGCCACTTAAGTCAACAGATTAAGGCGAGTGTGCAGTGTATGGATGTACGAC 3474
QY 3475 TACACCGCCCAAGACATACGAACTAGCTTTCAGCAAGGCCAGATCATCAACGTCTTC 3534
DB 3475 TACACCGCCCAAGATGACATGATGAGCTGGGCTTTCACCAAGGGCCAGATCATCAACGTCTTC 3534
QY 3535 AACCAAGAGAGACCCGAGCTGTGGAAGAGAGTACAGTGGGCACTTGGCTTCCCA 3594
DB 3535 AACCAAGAGAGACCTGTGAGTGTGGAAGAGAGTCAATGACAACTGGGCTTCCCA 3594
QY 3595 TCCAAATTATGTAAGCTGCACACAGCATGACCCCGACCAATGA 3642
DB 3595 TCCAAATTATGTAAGCTGCACACAGCATGACCCCGACCAATGA 3642
RESULT 9
ABK43498
ID ABK43498 standard; cDNA; 3319 BP.
AC ABK43498;
XX
DT 05-JUN-2002 (first entry)
XX
DE DNA encoding novel central nervous system protein #78.
XX
KW Central nervous system; CNS; autoimmune disease; rheumatoid arthritis;
KW hyperproliferative disorder; neoplasm; cardiovascular disorder;
KW cardiac arrest; cerebrovascular disorder; ischaemia; angiogenesis;
KW nervous system disorder; Alzheimer's disease; AIDS; ocular disorder;
KW acquired immunodeficiency virus; dysphagia; gastrointestinal disorder;
KW adenocarcinoma; reproductive system disorder; testicular feminisation;
KW endocrine disorder; diabetes; cancer; leukaemia; neovascularisation;
KW respiratory disorder; renal disorder; kidney failure; blood disorder;
KW myocardial infarction; wound healing; cell proliferation; skin aging;
KW food additive; food preservative; gene therapy; gene; ss.
OS Homo sapiens.
XX
XX WO200155318-A2.
FN
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01332.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.

PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239335.
PR 13-OCT-2000; 2000US-0239337.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0244674.
PR 08-NOV-2000; 2000US-0244675.
PR 08-NOV-2000; 2000US-0244676.
PR 08-NOV-2000; 2000US-0244677.
PR 08-NOV-2000; 2000US-0244678.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.

PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 DR MPI; 2001-581633/65.
 XX P-PSDB; AAU87168.
 PT
 PT New isolated nucleic acid encoding a protein for diagnosing,
 PT preventing, treating or ameliorating medical conditions and used as
 PT food additives or preservatives -
 PS
 XX Claim 1; SEQ ID No 88; 837pp; English.
 CC The invention describes an isolated nucleic acid molecule (I) encoding a
 CC novel central nervous system protein. (II) and polypeptides (III) encoded
 CC by (I), are used to treat a medical conditions and in diagnosis of a
 CC pathological condition. Disorders which are diagnosed or treated include
 CC autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
 CC disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
 CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
 CC angiodystrophies, nervous system disorders e.g. Alzheimer's disease and
 CC amyotrophic lateral sclerosis, infections caused by bacteria, viruses
 CC e.g. Acquired immunodeficiency virus (AIDS) and fungi, ocular disorders
 CC e.g. corneal infection, gastrointestinal disorders e.g. dysphagia,
 CC adenocarcinomas and irritable bowel syndrome, reproductive system
 CC disorders e.g. testicular feminisation, endocrine disorders e.g. diabetes
 CC and pituitary dwarfism, cancers and disorders at the cellular level e.g.
 CC leukaemia, disorders involving neovascularisation e.g. malignancies,
 CC respiratory disorders e.g. nonallergic rhinitis, renal disorders e.g.
 CC acute kidney failure and blood related disorders e.g. myocardial
 CC infection. The polypeptides can also be used to aid wound healing and
 CC epithelial cell proliferation, to prevent skin aging due to sunburn, to
 CC maintain organs before transplantation, for supporting cell culture of
 CC primary tissues, to regenerate tissues and in chemotaxis. The
 CC polypeptides can also be used as a food additive or preservative to
 CC increase or decrease storage capabilities, fat content, lipid, protein,

Query Match 56.1%; Score 2043.8; DB 23; Length 3319;
 Best Local Similarity 85.2%; Pred. No. 0;
 Matches 2324; Conservative 0; Mismatches 387; Indels 18; Gaps 3;
 QY 1 ATGGCTCAGTTTCCACACCTTTGCTGGTAGCTGTGAGCTGTGGCCATTAACCTGTGAG 60
 DB 159 ATGGCTCAGTTTCCACACCTTTGCTGGTAGCTGTGAGCTGTGGCCATTAACCTGTGAG 218
 QY 61 GAAAGGGCCAGCATGACCGAGCTTCCCTAGCTGAGAGCCGATGAGGGATTTACT 120
 DB 219 GAAAGAGCGAGCATGATGACGAGCTTCATAGTTTAAAGCCAAATCTGATTCATTA 278
 QY 121 GGTATCAACGAGGAACTTTTTCCTCAATCTGAGTTTACCTGAGCTGTCTTACGACAA 180
 DB 279 GGTATCAACGAGTAAACTTTTTCCTCAATCTGAGTTTACCTGAGCTGTCTTACGACAG 338
 QY 181 ATATGGGCGCTAGCGGACATGAATTAACATGAGAGAGATGATCAAGTGAATTTCCATA 240
 DB 339 ATATGGGCGCTAGCGGACATGAATTAACATGAGAGAGATGATCAAGTGAATTTCCATA 398
 QY 241 GCCATGAAGCTTATCAACTGAAGCTACAGAGATATCAGCTCCCTCCACACTTCCCT 300
 DB 399 GCTATGAACTTATCAACTGAAGCTACAGAGATATCAGCTCCCTCCACACTTCCCT 458
 QY 301 GTCATGAACAGCAACGAGCTGATTTCCAGTGCACACGATTTGGTATAGAGGAGATT 360
 DB 459 GTCATGAACAGCAACGAGCTGATTTCCAGTGCACACGATTTGGTATAGAGGAGATC 518
 QY 361 GCTAGATGCGACCACTGACAGCTGTTGCTCTGTGTCGAATGGGCTTCATTCAGTTGTT 420
 DB 519 GCGAGCATGCGACCGCTTACAGCTGTTGCTCTGTGTCGAATGGGCTTCATTCAGTTGTT 578
 QY 421 GGAATGCTCCACCTTATGATATCTTCTGCTCCACAGAGCTCCCTGGGCTTAAC 480
 DB 579 GGAATGCTCCACCTTATGATATCTTCTGCTCCACAGAGCTCCCTGGGCTTAAC 638
 QY 481 GGGGCTCCTCCCGTCATACAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
 DB 639 GGGGCTCCTCCCGTCATACAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 698
 QY 541 AAGAGTTCTTCTTACAGAGATCTGTGTCAGAGGTCAACATTAACATTAAGTACAGAG 600
 DB 699 AAGAGTTCTTCTTACAGAGATCTGTGTCAGAGGTCAACATTAACATTAAGTACAGAG 758
 QY 601 GCATATCATTCATATGTCGCGAGCGCCCTCCAGCAGCAAGATGGGCTGTGCTCACTCA 660
 DB 759 GCATATCATTCATATGTCGCGAGCGCCCTCCAGCAGCAAGATGGGCTGTGCTCACTCA 818
 QY 661 TCAAGGCTGAAATACAGGAGTATTCACAGCAGCAGCAAACTATGATGAGACACTTGA 720
 DB 819 TCAAGGCTGAAATACAGGAGTATTCACAGCAGCAGCAGCAAACTATGATGAGACACTTGA 878
 QY 721 AC-----AGGTCCCGAGGAGAACTATTTCTCATGCAATCAAGTTTACCCGAG 768
 DB 879 ACAGGTTCTGTATTAGGCTCCCAAGCAAGAACTATTTCTCATGCAATCAAGTTTACCCGAG 938
 QY 769 GCTAGAGCTGGTTAAATATGAAATCTTTGACATTTGATCAAGATGAAAACTACAGCA 828
 DB 939 GCTAGAGCTGGTTAAATATGAAATCTTTGACATTTGATCAAGATGAAAACTACAGCA 998
 QY 829 GAAAGATTTTCTAGGCTAGGACCTTAATGATGATGTCATGTCAGTGTGCTGAGCCACTGCG 888
 DB 999 GAAAGATTTTCTAGGCTAGGACCTTAATGATGATGTCATGTCAGTGTGCTGAGCCACTGCG 1058
 QY 889 CCCGTCGCTCCAGATATCATCTCTCTTCTTCAAGAGAGTTGCTCGGAGTGGG 948
 DB 1059 CCGTCGCTCCAGATATCATCTCTCTTCTTCAAGAGAGTTGCTCGGAGTGGG 1118
 QY 949 ATGTCGCTATTAAGCTCTTCTGTGATGATGAGGCTGCTGAGAGCCGCTCAAG 1008
 DB 1119 ATATCTGCTATTAAGCTCTTGTGATGATGAGGCTGCTGAGAGCCGCTCAAG 1178

QY 1009 GATGACGACAGC---CAGAGAGAACTGCTGTGACATTTGAGATAGAAGCCGAG 1065
 DB 1179 GATGAAACAACAATATAGAAAAAATTACCTGTAACTTTGATTAAGAAACCGGAG 1238
 QY 1066 AACTTTCAGCGAGCGAGTGTGTGAGCTGTGAGAACCGCCCGACAGCTTTTGAGCAGCAG 1125
 DB 1239 AACTTTGAACTGTGGCAACCTGTGAACTGTGAGAAACGAAAGGCAAGCTCTCTGGAAACAGCAG 1298
 QY 1126 CGCAAGACGAGGAGGTTGGCTCAGCTGTGAGCGGCGGAGAGAGAGAGAGAAAGCCG 1185
 DB 1299 CGCAAGAGAGAGAGCGCTGTGGCCAGCTGTGAGCGGCGGAGAGAGAGAGAGAGCGCT 1358
 QY 1186 GAGCGCCAGAGAGAGAGAGCGCAAGCCGAGCTGTGAGCTGTGAGAGAGAGAGAGAGAG 1245
 DB 1359 GAGCGCCAGAGAGAGAGAGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1418
 QY 1246 CGGAGAGCTGTGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1305
 DB 1419 CGGAGAGCTGTGAAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGCT 1478
 QY 1306 GCAAAACGAGAACTGTGAAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1365
 DB 1479 GCAAAACGAGAACTGTGAAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1538
 QY 1366 CTCCTGAAATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1425
 DB 1539 CTACTTAATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1598
 QY 1426 ACTCTGTGAGTTGAGTTAGAGAGCTGTGAATGACAAAAGAGATAGCTAGAGAGAGAG 1485
 DB 1599 ACTTTGGAATTTGAAATTAAGAGCTTAATTAATTAATTAATTAATTAATTAATTAAT 1658
 QY 1486 CAGGATATGAGGTGTGAGCTGTGCAACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1545
 DB 1659 CAAAGATATGAGATGTGATTTGACACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1718
 QY 1546 AGAGAGCTAAGATTTGCTGAATCACCACCTTACAGAGAGAGAGAGAGAGAGAGAGAG 1605
 DB 1719 AAGAGTTGAGATTTGCTGAATCACCACCTTACAGAGAGAGAGAGAGAGAGAGAGAG 1778
 QY 1606 ATGCTTGAAGAGCTTATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1665
 DB 1779 ATGCTTGAAGAGCTTATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1838
 QY 1666 CAGAAACGTTTGCATGAGAGAGCTGTCTTACCTTCAAAAAGAGCTTTGAGAGAGAG 1725
 DB 1839 CAGAAACGTTTGCATGAGAGAGCTGTCTTACCTTCAAAAAGAGCTTTGAGAGAGAG 1898
 QY 1726 CTGAGCCGAGAGAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1785
 DB 1899 CTAGCTCGGAGAGAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1958
 QY 1786 CTGAGAGAGATTTGATTTTCAACAAACAGCTGAGAGAGAGAGAGAGAGAGAGAGAG 1845
 DB 1959 CTACAGAGAGATTTGATTTTCAACAAACAGCTGAGAGAGAGAGAGAGAGAGAGAGAG 2018
 QY 1846 CAGCAACTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1905
 DB 2019 CAGCAACTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2078
 QY 1906 AGGAGAGAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1965
 DB 2079 CGAAAGATATAGATTTAGAAAAACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2138
 QY 1966 GACAAAGATGTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2022
 DB 2139 GACAAAGATGTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2198
 QY 2023 CAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2082
 DB 2199 CAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2258
 QY 2083 GCCAAGCCGAGAAATGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGCT 2142

DB 2259 GCGAAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2318
 QY 2143 AAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2202
 DB 2319 AAGCCAGCTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2378
 QY 2203 CAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2262
 DB 2379 CAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2438
 QY 2263 GATGAGATCAGCATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2322
 DB 2439 GATGAGATCAGCATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2498
 QY 2323 CAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2382
 DB 2499 CCGGCTGTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2558
 QY 2383 GAAAGATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2442
 DB 2559 GAGAAATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2618
 QY 2443 GCCCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2502
 DB 2619 GCCCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2678
 QY 2503 GAGCCCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2562
 DB 2679 GAGCCCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2738
 QY 2563 AATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2622
 DB 2739 AATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2798
 QY 2623 CTAAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2682
 DB 2799 CTAAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2858
 QY 2683 TCCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2711
 DB 2859 TCCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2887

RESULT 10
 AAI63825
 ID AAI63825 standard; cDNA, 3466 BP.
 XX
 AC AAI63825;
 XX
 DT 22-OCT-2001 (first entry)
 XX
 DE Human polynucleotide SEQ ID NO 33.
 XX
 OS Homo sapiens.
 OS
 PN WO200155308-A2.
 PN
 PD 02-AUG-2001.
 PD
 XX
 PF 17-JAN-2001; 2001WO-US01309.
 PF
 PR 31-JAN-2000; 2000US-0179065.
 PR
 PR 04-FEB-2000; 2000US-0180628.
 PR
 PR 24-FEB-2000; 2000US-0184664.

CC The invention relates to human polynucleotides (A163803-A164012) and
 CC the encoded proteins (A443497-A443660) useful for preventing, treating
 CC or ameliorating medical conditions e.g. by protein or gene therapy. The
 CC genes were isolated from a range of human tissues disclosed in the
 CC specification. The nucleic acids, proteins, antibodies and (ant)agonists
 CC are useful in the diagnosis, treatment and prevention of: (a) cancer,
 CC e.g. breast and ovarian cancer and other cancers of the adrenal gland,
 CC bone marrow, breast, gastrointestinal tract, liver, lung, or
 CC urogenital; (b) immune disorders e.g. Addison's disease, allergies;
 CC autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus,
 CC Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
 CC and parasitic infections.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.

CC Sequence 3466 BP; 1056 A; 829 C; 861 G; 710 T; 10 other;

CC Query Match 56.1%; Score 2043.8; DB 22; Length 3466;

CC Best Local Similarity 85.2%; Pred. No. 0;

CC Matches 2324; Conservative 0; Mismatches 387; Indels 18; Gaps 3;

QY 1 ATGGCTCAGTTTCCACACCTTGGGTAGCTGATGTCGGGCATTAACCTGAGG 60
 DB 277 ATGGCTCAGTTTCCACACCTTGGGTAGCTGATGTCGGGCATTAACCTGAGG 336
 QY 61 GAAAGGCGCAAGCATGATGATGATGATGATGATGATGATGATGATGATGAT 120
 DB 337 GAAAGGCGCAAGCATGATGATGATGATGATGATGATGATGATGATGATGAT 396
 QY 121 GGTGATCAAGCCAGGAACTTTTTCATCTGGGTACCTGAGCTGCTTGAAGCAA 180
 DB 397 GGTGATCAAGCCAGGAACTTTTTCATCTGGGTACCTGAGCTGCTTGAAGCAA 456
 QY 181 ATATGGGCGCTAGCGGACATGATGATGATGATGATGATGATGATGATGATGAT 240
 DB 457 ATATGGGCGCTAGCGGACATGATGATGATGATGATGATGATGATGATGATGAT 516
 QY 241 GGCATGAACCTTATCAAACTGAAAGTACAAAGATATCACTCCCTCCACACTTCC 300
 DB 517 GGCATGAACCTTATCAAACTGAAAGTACAAAGATATCACTCCCTCCACACTTCC 576
 QY 301 GTCATGAACCAAGCAACCACTGCTATTTTCACTGCAACGACATTTGGTATGAG 360
 DB 577 GTCATGAACCAAGCAACCACTGCTATTTTCACTGCAACGACATTTGGTATGAG 636
 QY 361 GGTAGCATGCAACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
 DB 637 GGTAGCATGCAACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 696
 QY 421 GGAATGCTCCACCTTATGATCTTCTGCTCCCTCCAGCAGAGTGCCTCCCTGCTAAC 480
 DB 697 GGAATGCTCCACCTTATGATCTTCTGCTCCAGCAGAGTGCCTCCCTGCTAAC 756
 QY 481 GGGGCTCTCCGCTCATAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
 DB 757 GGGGCTCTCCGCTCATAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 816
 QY 541 AAGAGTCTTCTCTGAGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600
 DB 817 AAGAGTCTTCTCTGAGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 876
 QY 601 GCAATATCATTCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
 DB 877 GCAATATCATTCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 936
 QY 661 TCAGGCTGAATATGAGGCTTATTCACAGCAGCAAACTATGAGTGACCTTA 720
 DB 937 TCAGGCTGAATATGAGGCTTATTCACAGCAGCAAACTATGAGTGACCTTA 996

QY 721 AC-----AGTCCCGAGGCAAGCAATATTCATGCAATCAAGTTTACCCGAG 768
 DB 997 AAGAGTCTCTTGTAGTGTCCAGCAGAACTATTTTATGACAGCAAGTTTACCA 1056
 QY 769 GCTCAGCTGCTTCAATATGAAATCTTTCTGACATTTGATCAAGATGAAACTCACTGCA 828
 DB 1057 GCTCAGCTGCTTCAATATGAAATCTTTCTGACATTTGATCAAGATGAAACTCACTGCA 1116
 QY 829 GAAAGATTTATCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 888
 DB 1117 GAAAGATTTATCTGCAATGACCTCATATGATGATGATGATGATGATGATGATGAT 1176
 QY 889 CCCGCTCTCTCCAGAAATACATCCCTCTCTCAGAAAGTTCGCTCCGCACTGCG 948
 DB 1177 CCGCTCTCTCTCCAGAAATACATCCCTCTCTCAGAAAGTTCGCTCCGCACTGCG 1236
 QY 949 ATGTCGTCATAGCTCTTCTTCTGATGATGATGATGATGATGATGATGATGATGAT 1008
 DB 1237 ATATCTGTCATAGCTCAATCTGATGATGATGATGATGATGATGATGATGATGAT 1296
 QY 1009 GATGACGACGAC---CAGAGAAAGAACTGCTGTCATTTGAAGATTAAGACCGGAG 1065
 DB 1297 GATGAAACAACAATTAAGAAATTAATTAATTAATTAATTAATTAATTAATTAAT 1356
 QY 1066 AACTTCAGAGCAGGAGCTGATGATGATGATGATGATGATGATGATGATGATGAT 1125
 DB 1357 AACTTCAGAGCAGGAGCTGATGATGATGATGATGATGATGATGATGATGATGAT 1416
 QY 1126 CGCAAGAGCAGAGCGGCTGCTCACTGAGCGCGCCGAGCAGAGAGAGAAAGCCG 1185
 DB 1417 CGCAAGAGCAGAGCGGCTGCTCACTGAGCGCGCCGAGCAGAGAGAGAAAGCCG 1476
 QY 1186 GAGCCGAGAGCAGAGAGCGCAGAGCGCAGCTGAGCTGAGAGAGAGAGAGAG 1245
 DB 1477 GAGCCGAGAGCAGAGAGCGCAGAGCGCAGCTGAGAGAGAGAGAGAGAGAGAG 1536
 QY 1246 CGGAGCTGAGCAGGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1305
 DB 1537 CGGAGCTGAGCAGGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1596
 QY 1306 GCAAAAACGGGAACTGGAAGAGCAGCACTTGAATGGGAAACGGAACCGAGACAG 1365
 DB 1597 GCAAAAACGGGAACTGGAAGAGCAGCACTTGAATGGGAAACGGAATCGAGAG 1656
 QY 1366 CTCTGATCAGAGGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1425
 DB 1657 CTACTTAATCAAAAGAAACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1716
 QY 1426 ACTCTGAGTTGAGTTGAAGCTGTAATGACCAAAAGCATCAGTGAAGAGAAACCT 1485
 DB 1717 ACTCTGAGTTGAGTTGAAGCTGTAATGACCAAAAGCATCAGTGAAGAGAAACCT 1776
 QY 1486 CAGATATCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1545
 DB 1777 CAGATATCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1836
 QY 1546 AAGAGCTGAAGATTTGCTGAATTCACCTTAACGACAGCACTGAGAGATCTAC 1605
 DB 1837 AAGAGCTGAAGATTTGCTGAATTTCAACCTTAACGAGCACTGAGAGATCTAC 1896
 QY 1606 ATGCTTGAAGCTTATTCAGAGAAAGAGATCAGTGAAGAGAGAGAGAGAGAG 1665
 DB 1897 ATGCTTGAAGCTTATTCAGAGAAAGAGATCAGTGAAGAGAGAGAGAGAGAG 1956
 QY 1666 CAGAACAGTTGCTATGAGACTGCTTCTTCACTCAAAAGAGCTTGAAGCAAGAG 1725
 DB 1957 CAGAACAGTTGCTATGAGAGATTTCACTTTTACCTTAAAGAGCTTGAAGCAAGAG 2016
 QY 1726 CTGGCCCGGAGAGAGCTCGGAGAGAGCTGAGCAGAGTGAAGAGAGAGAGAG 1785
 DB 2017 CTAGCTCGGAGAGAGCTTCAAGAGCACTGAGTGAAGTGAAGAGAGAGAGAG 2076
 QY 1786 CTGAGAGAGATGATGATTTTCAACCAACAGCTGAAGAGAGATGAGAGAGATGAG 1845

```

Db      2077  CTACAGGAGATTGATTTTCAATTAATCAGCTGAAGAGAACTAAGAAATACACATATAG 2136
Qy      1846  CAGCAATCTCCAGAACGAGAGTCCCTGAGAGCAGCCCACTGAAAGCAGAAAGAGAGAG 1905
Db      2137  CAACAATCTCCAGAACGAGAGTCCCTGAGAGCAGCCCACTGAAAGCAGAAAGAGAGAG 2196
Qy      1906  AGSAAAGACCTGGAGTTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1965
Db      2197  CGAAAGATCATATGAAATTAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2256
Qy      1966  GACAAGCAATGGCTGGAGCATGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2022
Db      2257  GACAAGCAATGGCTGGAGCATGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2316
Qy      2023  CACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2082
Db      2317  CACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2376
Qy      2083  GCCAAGCCGGAAATGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2142
Db      2377  GGCAAAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2436
Qy      2143  AAGCTGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2202
Db      2437  AAGCAGAGCTGTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2496
Qy      2203  CAGGAGAGTGAATGATGATATATATATATATATATATATATATATATATATATATAT 2262
Db      2497  CAGGAGAGTGAATGATGATATATATATATATATATATATATATATATATATATATAT 2556
Qy      2263  GATGAGATCAGATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2322
Db      2557  GATGAGATCAGATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2616
Qy      2323  CCAGGATGAGCTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2382
Db      2617  CCGGCTGCTGGCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2676
Qy      2383  GAAAGAGTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2442
Db      2677  GAGAGAGATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2736
Qy      2443  GCGGCTGCGCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2502
Db      2737  GCGGCTGCGCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2796
Qy      2503  GAGGCTGCGCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2562
Db      2797  GAGGCTGCGCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2856
Qy      2563  AACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2622
Db      2857  AATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2916
Qy      2623  CCTAGTGTGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2682
Db      2917  CCAAGTGTGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2976
Qy      2683  TCCCATCTCCGCTGCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2711
Db      2977  TCCCGTCTCTCTGTCTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3005

```

RESULT 11
AAZ34573
ID AAZ34573 standard; cDNA; 2079 BP.

AC AAZ34573;
XX 01-FEB-2000 (first entry)
XX Human SH3D1A cDNA clone 5.

```

XX      SH3D1A gene: human; Down's syndrome; Leukemia; cancer;
KW      megakaryocytic abnormality; myeloproliferative disorder;
KW      platelet disorder; neural disorder; thrombocytopenia;
KW      haematopoietic disorder; cognitive dysfunction; microcephaly;
KW      lissencephaly; colpocephaly; holoprosencephaly; diagnosis; therapy;
KW      ss.
XX      Homo sapiens.
XX      Key Location/Qualifiers
FH      Key 136..2079
FT      CDS /*tag= a
FT      MO9953062-A2.
PD      21-OCT-1999.
PF      16-APR-1999; 99MO-US08371.
PE      16-APR-1998; 98US-0082007.
PR      (CEDA-) CEDARS SINAI HEALTH SYSTEM.
PA      Korenberg JR, Chen X;
PI      WPI: 1999-633829/54.
DR      P-PSDB; AAY32157.
PT      Nucleic acid from the human SH3D1A gene and its products, useful for
PT      the diagnosis and treatment of myeloproliferative disorders and
PT      leukemia -
PS      Claim 2; Fig 12; 99pp, English.
XX      This is the nucleotide sequence of a cDNA clone, termed clone 5,
XX      corresponding to a novel human SH3 gene, termed the SH3D1A gene,
XX      that contributes to the development of platelets and the
XX      pathogenesis of leukemias, both in general and in particular those
XX      involving the megakaryocytic lineage. The SH3D1A gene maps to the
XX      small candidate region for low platelets on chromosome 21.
XX      Sequencing of 5 different sizes of cDNA clone from foetal brain
XX      (see AAZ34570-74) suggests that at least 3 isoforms exist. The
XX      invention provides methods for the diagnosis and treatment of
XX      megakaryocytic abnormality, myeloproliferative disorder, platelet
XX      disorder, acute leukaemia, neural disorders, thrombocytopenia,
XX      platelet disorder on chromosome 21, low platelets in deletion for
XX      21, association of gains in chromosome 21 with leukemias, neural
XX      abnormalities, dysfunctions and disorders including brain
XX      malformations and corresponding cognitive dysfunctions,
XX      microcephaly, lissencephaly, and colpocephaly. Methods are also
XX      provided for: suppressing cells unable to regulate themselves;
XX      screening for a somatic alteration in the SH3D1A gene; monitoring
XX      the progress and adequacy of a treatment; monitoring tumour risk
XX      progress or megakaryocytic abnormality, myeloproliferative disorder,
XX      haematopoietic disorder, platelet disorder or leukemias; and
XX      treatment of a subject (including a prenatal subject) having
XX      megakaryocytic abnormality, myeloproliferative disorder,
XX      platelet disorder, leukaemia or neural disorder using a
XX      nucleic acid that expresses SH3D1A or its antisense nucleic acid.
XX      Sequence 2079 BP; 700 A; 451 C; 495 G; 433 T; 0 other;
SQ

```

Query Match 41.0%; Score 1495; DB 20; Length 2079;
Best Local Similarity 85.9%; Pred. No. 0;
Matches 1670; Conservative 0; Mismatches 270; Indels 3; Gaps 1;

```

Qy      1  ATGGCTCAGTTTCCACACCTTTCGGTGTAGCCTGATGCTGGCGCATTAACGTGGAG 60
Db      136  ATGGCTCAGTTTCCACACCTTTCGGTGTAGCCTGATGCTGGCGCATTAACGTGGAG 195
Qy      61  GAAAGGCCAAGCATGAGCAGAGTTCCTTAGCCTGAAGCCGATAGCGGATTTAACT 120

```

Db 196 GAAAGACGAGCATGATGACAGTTTCATAGTTTAAAGCAATATCTGATTCATTACT 255
 QY 121 GGTGATCAAGCGAGAACTTTTTCATCTGGGTTTCTCAGCTGTCTTAGACAA 180
 Db 256 GGTGATCAAGCTTGAATCTTTTTCATCTGGGTTTCTCAGCTGTCTTAGACAA 315
 QY 181 ATATGGGCGCTAGCGGACATGATTAACAATGGAAGATGATCAATGGAATTTTCCATA 240
 Db 316 ATATGGGCGCTAGCTGATGATTAACAATGGAAGATGATCAATGGAATTTTCCATA 375
 QY 241 GCGATGAAGCTTTCATCACTGAAAGCTAATGAGTCCCTCCCACTTCCCT 300
 Db 376 GGTATGAATCTTATCAACTGAAGCTAATGAGTCCCTCCCACTTCCCT 435
 QY 301 GTCATGAACAGCAACAGTGGCTATTTCCAGTGCACGAGCTTGTGTAGAGGAGAT 360
 Db 436 GTCTATGAACAGCAACAGTGGCTATTTCCAGTGCACGAGCTTGTGTAGAGGAGAT 495
 QY 361 GGTAGCATGCGACCACTCAGCTGTGCTCTGCTCCCAATGGGCTCCATTCAGTTGTT 420
 Db 496 GCGAGATGCGACCGCTTACAGCTGTGCTCCAGTGCACATGGGATCCATTCAGTTGTT 555
 QY 421 GGAATGCTTCACACCTTAGATCTTCTGCTCCCAAGAGAGTGGCTCCCTGGCTAAC 480
 Db 556 GGAATGCTTCACACCTTAGATCTTCTGCTCCCAAGAGAGTGGCTCCCTGGCTAAC 615
 QY 481 GGGGCTCCCTCCGCTCAATAGCTCTGCTGGCTTGGCTTGGCTTGGCTTGGCTTGGCT 540
 Db 616 GGGGCTCCCTCCGCTCAATAGCTCTGCTGGCTTGGCTTGGCTTGGCTTGGCTTGGCT 675
 QY 541 AAGAGTCTTCCCTTCCAGAGATCTGCTCAGAGGCTCAATTAACAATACTAAGTTACAGAG 600
 Db 676 AAGAGTCTTCCCTTCCAGAGATCTGCTCAGAGGCTCAATTAACAATACTAAGTTACAGAG 735
 QY 601 GCAATATCATTTGATGATGCTGCGCAGCGCTCCCAAGAGAGAGTGGCTTGGCTTGGCT 660
 Db 736 GCAATATCATTTGATGATGCTGCGCAGCGCTCCCAAGAGAGAGTGGCTTGGCTTGGCT 795
 QY 661 TCAAGGCTGAATATCAAGGAGTTATTCACAGGCAAGCAAACTTATGATGAGCACTTA 720
 Db 796 TCAAGGCTGAATATCAAGGAGTTATTCACAGGCAAGCAAACTTATGATGAGCACTTA 855
 QY 721 AAGAGTCTTCCCTTCCAGAGATCTGCTCAGAGGCTCAATTAACAATACTAAGTTACAGAG 780
 Db 856 AAGAGTCTTCCCTTCCAGAGATCTGCTCAGAGGCTCAATTAACAATACTAAGTTACAGAG 915
 QY 781 TCAATATGGAATCTTTCATGACATTTGAATGAATGAATGAATGAATGAATGAATGAAT 840
 Db 916 TCAATATGGAATCTTTCATGACATTTGAATGAATGAATGAATGAATGAATGAATGAAT 975
 QY 841 CTAGCTATGCACTTAATTTGATGTTGCAATGCTGCTGCTGCTGCTGCTGCTGCTGCT 900
 Db 976 CTAGCTATGCACTTAATTTGATGTTGCAATGCTGCTGCTGCTGCTGCTGCTGCTGCT 1035
 QY 901 CCGAATATGATCCCTCTCTTCCAGAGATTTGCTCCGCGAGTGGAGTGTCCGTCATA 960
 Db 1036 CCGAATATGATCCCTCTCTTCCAGAGATTTGCTCCGCGAGTGGAGTGTCCGTCATA 1095
 QY 961 AGCTCTTCTTCTGCTGATGAGAGCTGCTGAGAGCCGCTGCTGAGAGATGAGCAGCAG 1020
 Db 1096 AGCTCTTCTTCTGCTGATGAGAGCTGCTGAGAGCCGCTGCTGAGAGATGAGCAGCAG 1155
 QY 1021 C---CAGAGAGAGAACTGCTGCTGAGATTTGAAGATTAAGAGCGGAGAACTTCCAGGCA 1077
 Db 1156 CAATTATGAAGAAATTTACCTGTAAGATTTGAAGATTAAGAGCGGAGAACTTGAACCT 1215
 QY 1078 GCGAGTGTGAGCTGAGAGAGCGCGCAGAGCGCTTCTTGAAGAGAGAGCGCGCAAGAGCAG 1137
 Db 1216 GCGAGTGTGAGCTGAGAGAGCGCGCAGAGCGCTTCTTGAAGAGAGAGCGCGCAAGAGCAG 1275
 QY 1138 GAGCGGTTGGCTTCACTGAGAGCGCGCAGAGCGAGAGAGAGAGAGAGAGAGAGAGAGAG 1197
 Db 1276 GAGCGGTTGGCTTCACTGAGAGCGCGCAGAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1335

QY 1198 CAGAGGCGCAAGCGGAGCTGAGAGTGGAGAGAGCAGCTGGAGAGAGAGAGAGAGAGAG 1257
 Db 1336 CAGAGGCGCAAGCGGAGCTGAGAGTGGAGAGAGCAGCTGGAGAGAGAGAGAGAGAGAG 1395
 QY 1258 CCGCAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1317
 Db 1396 CCGCAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1455
 QY 1318 CTGGAAGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1377
 Db 1456 CTGGAAGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1515
 QY 1378 AGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1437
 Db 1516 AGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1575
 QY 1438 GAGTTAGAGAGCTGTAATGAGCAAAAGAGATGAGTGAAGAGAGAGAGAGAGAGAGAGAG 1497
 Db 1576 GAGTTAGAGAGCTGTAATGAGCAAAAGAGATGAGTGAAGAGAGAGAGAGAGAGAGAGAG 1635
 QY 1498 TGTGCACTGGCAACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1557
 Db 1636 TGTGCACTGGCAACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1695
 QY 1558 ATTGCTGAAATGACCCACTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1617
 Db 1696 ATTGCTGAAATGACCCACTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1755
 QY 1618 CTTATTTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1677
 Db 1756 CTTATTTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1815
 QY 1678 CATAGAGAGCTGCTTCTTACCTTAATAAGAGCTTGAAGAGAGAGAGAGAGAGAGAGAGAG 1737
 Db 1816 CATAGAGAGCTGCTTCTTACCTTAATAAGAGCTTGAAGAGAGAGAGAGAGAGAGAGAGAG 1875
 QY 1738 CAGCTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1797
 Db 1876 CAGCTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1935
 QY 1798 GATGTTTTCACACACAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1857
 Db 1936 GATGTTTTCACACACAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1995
 QY 1858 AAGCAGAGGCTCCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1917
 Db 1996 AAGCAGAGGCTCCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2055
 QY 1918 CAGTTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1940
 Db 2056 CAGTTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2078

RESULT 12
 AAK94139
 ID AAK94139 standard; cDNA, 2131 BP.

AC AAK94139;
 AC 06-NOV-2001 (first entry)
 DE Human full-length cDNA, SEQ ID NO: 2646.
 XX Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
 XX Homo sapiens.
 OS Homo sapiens.
 KW Homo sapiens.
 PN EPI130094-A2.
 XX
 XX 05-SEP-2001.
 PD
 XX
 PF 07-JUL-2000; 2000EP-0114089.

1825 CTTATTCAGAAAACAGATCACTGACCAATTAACCACTTCAGACGAAACAGTTTG 1884
1678 CATAGAGACTCGGTTCTTACCTCCAAAGAGCCTTGAGACCAAGAGCTGCCCGGCGAG 1737
1885 CACAGAGATTCACTTTTACCTTAAAGAGCCTTGAGACCAAGAGCTGCCCGGCGAG 1944
1738 CAGCTCCGGGAGCAGCTGACGAGTGTGAGAGAGAGCCAGGCTCAAGCTCGAGAGATT 1797
1945 CACCTACGAGACCAACTGATGTAAGTGGAGAAAGAACTGATCAAACTACAGAGAGATT 2004
1798 GATGTTTTCAACACCAAGCTGAAAGAACTGAGAGATACATAGCAAAACGACACTCCAG 1857
2005 GATATTTTCAATATATCAGCTGAAAGAACTGAAAGAAATACACAATTAAGCAACAACTCCAG 2064
1858 AAGCAAGAGTCTCTGAGAGCAGCGGCACTGTAACAGAAAGAGAGAGAGAGAGCTTG 1917
2065 AAGCAAAATTCATGTGAGGCTGAACGACTGAAACAGAAAGAACAGAAAGATCATTA 2124
1918 GAGTTAG 1924
2125 GAATTAG 2131

RESULT 13
AAS84762
ID AAS84762 standard; cDNA; 2874 BP.
AAS84762;
13-FEB-2002 (first entry)
DNA encoding novel human diagnostic protein #20566.
Human; chromosome mapping; gene mapping; gene therapy; forensic;
food supplement; medical imaging; diagnostic; genetic disorder; ss.
Homo sapiens.
WO200175067-A2.
11-OCT-2001.
30-MAR-2001; 2001WO-US08631.
31-MAR-2000; 2000US-0540217.
23-AUG-2000; 2000US-0649167.
(HYSE-) HYSBO INC.
Dymanac RT, Liu C, Tang YT;
WPI; 2001-639362/73.
P-PSDB; ABG20575.
New isolated polynucleotide and encoded polypeptides, useful in
diagnostics, forensics, gene mapping, identification of mutations
responsible for genetic disorders or other traits and to assess
biodiversity -
Claim 1; SEQ ID No 20566; 103bp; English.

The invention relates to isolated polynucleotide (I) and
polypeptide (II) sequences. (I) is useful as hybridisation probes,
polynucleotide chain reaction (PCR) primers, oligomers, and for chromosome
and gene mapping, and in recombinant production of (II). The
polynucleotides are also used in diagnostics as expressed sequence tags
for identifying expressed genes. (I) is useful in gene therapy techniques
to restore normal activity of (II) or to treat disease states involving
(II). (II) is useful for generating antibodies against it, detecting or
quantifying a polypeptide in tissue, as molecular weight markers and as
a food supplement. (II) and its binding partners are useful in medical
imaging of sites expressing (II). (I) and (II) are useful for treating

disorders involving aberrant protein expression or biological activity.
The polypeptide and polynucleotide sequences have applications in
diagnostics, forensics, gene mapping, identification of mutations
responsible for genetic disorders or other traits to assess biodiversity
and to produce other types of data and products dependent on DNA and
amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2874 BP; 940 A; 679 C; 703 G; 552 T; 0 other;
Query Match 35.5%; Score 1291.4; DB 23; Length 2874;
Best Local Similarity 76.0%; Pred. No. 2.1e-298;
Matches 1778; Conservative 0; Mismatches 266; Indels 297; Gaps 3;
QY 1303 GCCGCAAAACGGAACTGGAAAGCGACGACCACTTGAATGGGAACGAAACCGAGACAG 1362
1 GCTGCAAAACGGAACTGGAAAGCGACGACCACTTGAATGGGAACGAAACCGAGCA 60
QY 1363 GAACTCTGGAATCAAGAGAAACAAGAGAGAGGACCGTGCTCTGAAGCCAGAGAG 1422
61 GAACTACTAATCAAAAGAAACAAGAGAGACATAGTTGTAAGCAAAAGAA 120
QY 1423 AAGACTCTGAGATTGAGTTAGTAAGCTCTGAATGACAAAACCTAGCTAAGAGAAA 1482
121 AAGACTTGGAAATTTGAATTAAGCTCTAATGATTAAGATCAATCACTAAGAGGAAA 180
QY 1483 CTTACAGATATCAGGCTGCTGACCTGGCAACCGAGAGCAAGAAATTGAGAGCAAGAAC 1542
181 CTTCAAGATATCAGATGTGATTTGACCAACCCAAAGCAAAATTGAGGCACAAACAA 240
QY 1543 TCTAGAGACTAAGAAATTTGCTGAATCACCCTTAACAGACAGATTGAGAGAAATCG 1602
241 TCTAGAGAGTTGAGAAATTCGCAAAATCACCCTTAACAGCAAAATTAACAGAAATTCG 300
QY 1603 CAAATGCTTGGAAAGACTTTTCCAGAGAAACAGATCTAGTGCACGATTAACAAAGTC 1662
301 CAAATGCTTGGAAAGACTTTTCCAGAGAAACAGATCTAGTGCACGATTAACAAAGTC 360
QY 1663 CAGCAGAACAGTTTGATGATGAGACTCGCTTCTTACCCCTCAAAAGAGCTTGAAGCAAG 1722
361 CAGCAGAACAGTTTGATGATGAGACTCGCTTCTTACCCCTTGAAGCAAG 420
QY 1723 GAGCTGGCCCGCAGACAGCTCCGGAGCACTGACAGAGTGTGAGAGAGACCAAGTCA 1782
421 GAACTAGCTCGGCAAGACCTTACGAGACCAACTGATGAAAGTGAAGAAAGAACTGATCA 480
QY 1783 AAGCTGCAAGAGATTGATTTTCAACAAACGCTGAGAGAACTGAGAGATTAATAC 1842
481 AAGCTGCAAGAGATTGATTTTCAACAAACGCTGAGAGAACTGAGAGATTAATAC 540
QY 1843 AAACAGCAACTCCAGAGAGAGAGTCCCTGAGGAGAGCGCGACTGAGAGAGAAAGCGAG 1902
541 AAGCAACAATCTCAGAGAGAGAAAGTCCATGAGGCTGAAAGCTGAAAGAGAGAGAGAA 600
QY 1903 GAGAGAGAGAGCTGAGATTGAGAGAGCAAAAGAGAGCTCAAGAGAGAGATTGAGAA 1962
601 GAGAGAGAGAGCTGAGATTGAGAGAGCAAAAGAGAGAGAGAGAGAGAGAGAGAGAG 660
QY 1963 AAGGACAGCAATGCTGAGAGAGATGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2019
661 AAGGACAGCAATGCTGAGAGAGATGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
QY 2020 CCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2079
721 CTCACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
QY 2080 AAGAGCAAGCCGAAATGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2139
781 AAGAGCAAGCCGAAATGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840

2140 GCTAAGTGGCCACCCAGGACCCCTGCTCTACCAAGAGAAAGCCCGCTTACATTTCT 2199
 841 GCTAAGCCAGCTGTCTCAGGCACTCTGTCTCAGTGAAGAAAGGCTCCATTTCT 900
 2200 GCACAGAGAGTGTAAAGTGTATATTACCGAGCGCTGTACCCCTTTGAAATCCAGAAAT 2259
 901 GCACAGAGAAATGTAAAGTGTATATTACCGAGCGCTGTACCCCTTTGAAATCCAGAAAT 960
 2280 CACATGTGATTCACATTCAGCAGAGATATATGATGTGTGTGTGAAAGCCAGCTGGA 2319
 961 CATGTGTAATCATATCTCAGCAGAGATATATGATGTGTGTGAAAGCCAGCTGGA 1020
 2320 GAGCCAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2379
 1021 GAACCCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1080
 2380 GCAGAAAGATTCAGAAATGAGGTTCCATCTCAGCCAAACAGTGAACGATCTGACA 2439
 1081 GCAGAGAAATTCAGAAATGAGGTTCCATCTCAGCCAAACAGTGAACGATCTGACA 1140
 2440 TCTGCCCCCTGCCCCCACTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2499
 1141 TCTGCCCCCTGCCCCCACTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1200
 2500 TCTGAGCCCTCCACAACCCCAACAACCTGAGGAGACTTCACTTCCAGTGGCCAGCAG 2559
 1201 TCAAGAGCCCTCCACAACCCCAACAACCTGAGGAGACTTCACTTCCAGTGGCCAGCAG 1260
 2560 TCMAACGAAAGCCAGAAACCGACAACTGGGATACGTGGGCGGCTCACTTCTGTAGC 2619
 1261 ACGAATGAGAAACCGAAACCGATATCTGGGATGATGGGAGCCAGCCCTCTCTCAC 1320
 2620 GTACTATGCTGTGGCAATTAACGAGAGATCAAGCTTACCCCGACCAAGCCAGCTGGC 2679
 1321 GTTCCAAATGCGCGCACTTAAGGAGAGTCCGCTTATCTCCAGCCAGCCAGCTGGC 1380
 2680 TCTCTCCCATCTCCCGCTCTGGGCGAGGAGTGAAGAGGCTCAAGAGCCAGCAGC 2739
 1381 TCTCTCCCGCTCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1440
 2740 CTGTATCCCTGTGAGAGCCAAAGAGCAACCACTTAAATTTTAAACAAAGTGAAGCTATC 2799
 1441 CTATATCTCTGTGAGAGCCAAAGAGCAACCACTTAAATTTTAAACAAAGTGAAGCTATC 1500
 2800 ACCGTTCTGGAACACAGACATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2859
 1501 ACCGTTCTGGAACACAGACATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1543
 2860 TTCCCAAGTCTTACGTAAACTCATTTCAAGGCGCGTAAGGAATCCACAAGCATCAT 2919
 1544 ----- 1543
 2920 ACTGCGCTACTGAAAGTCTGTAGTCTAAAGAGAGTGGCTTCCCGCGCCGCAAGCCA 2979
 1544 ----- 1543
 2980 GCCATTCCCGAGAGATTTATTTGCAATGACATACAGAGATTTCTGAGCAAGAGAT 3039
 1544 ----- 1543
 3040 TTAACCTTTCAGCAAGGAGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3099
 1591 TTAACCTTTCAGCAAGGAGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1650
 3100 GGAACGTTGGGAGCAAGTCCGAGTCTTCCCTTAACTATGTGAGGCTTAAAGATTCA 3159
 1651 GGAACGTTGGGAGCAAGTCCGAGTCTTCCCTTAACTATGTGAGGCTTAAAGATTCA 1710
 3160 GAGGCTCTGGAACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3219
 1711 GAGGCTCTGGAACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1770
 3220 GTTATTTGCTTCTTACGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3279

1771 GTTATTTGCTTCTTACGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1830
 3280 ATTCTGATCCGAGAAAGAAACCCAGGTGATGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3339
 1831 ATTCTGATCCGAGAAAGAAACCCAGGTGATGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1852
 3340 AAAAAGGCCAGATAGGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3399
 1853 ----- 1852
 3400 AGCAAAATCAACCCCACTGAGCTACCAAGACCGCAGTGCAGCAGAGTGTGCAGGTG 3459
 1853 ----- 1852
 3460 ATCGGATGTACGATTTACCTGCGCCCAAGACGATGACCACTAGCTTCAAGAAAGCCAG 3519
 1867 ATTGGATGTACGATTTACCTGCGCCCAAGACGATGACCACTAGCTTCAAGAAAGCCAG 1926
 3520 ATCATCAACGTCCTCAACCAAGAGAGCCCGGACTGTGTGTGTGTGTGTGTGTGTGTGTGT 3579
 1927 ATCATCAACGTCCTCAACCAAGAGAGCCCGGACTGTGTGTGTGTGTGTGTGTGTGTGTGT 1986
 3580 GTTGGGCTCTTCCCATCCAAATTATGTAAAGCTGACCAAGACATGACCCAGCAGCAA 3639
 1987 GTTGGGCTCTTCCCATCCAAATTATGTAAAGCTGACCAAGACATGACCCAGCAGCAA 2046
 3640 T 3640
 2047 T 2047

RESULT 14
 AAH1578
 ID AAH1578 standard; cDNA; 2131 BP.
 XX AAH1578;
 AC
 XX
 XX
 XX 26-JUN-2001 (first entry)
 DE Human cDNA sequence SEQ ID NO:15658.
 XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
 XX Homo sapiens.
 XX OS
 XX EP1074617-A2.
 XX PD
 XX 07-FEB-2001.
 XX
 XX 28-JUN-2000; 2000EP-0116126.
 XX
 XX 29-JUL-1999; 99JP-0248036.
 XX 27-AUG-1999; 99JP-0300253.
 XX 11-JAN-2000; 2000JP-0118776.
 XX 02-MAY-2000; 2000JP-0183767.
 XX 09-JUN-2000; 2000JP-0241899.
 XX
 XX (HELI-) HELIX RES INST.
 XX
 XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saio K, Yamamoto J;
 XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX WPI; 2001-318749/34.
 XX
 XX Primer sets for synthesizing polynucleotides, particularly the 5602
 XX full-length cDNAs defined in the specification, and for the detection
 XX and/or diagnosis of the abnormality of the proteins encoded by the
 XX full-length cDNAs -
 XX Claim 8; SEQ ID 15658; 2537bp + CD ROM; English.
 XX
 XX The present invention describes primer sets for synthesizing 5602

CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to a
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialized methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
 CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.

XX Sequence 2131 BP; 691 A; 448 C; 572 G; 420 T; 0 other;

Query Match 34.6%; Score 1259.2; DB 22; Length 2131;
 Best Local Similarity 81.2%; Pred. No. 9,1e-291;
 Matches 1339; Conservative 0; Mismatches 243; Indels 114; Gaps 2;

QY 1 ATGGCTCAGTTTCCACACCTTTCGCTGAGCTGATGTCGAGCCATACTGTGAG 60
 Db 347 ATGGCTCAGTTTCCACACCTTTCGCTGAGCTGATGTCGAGCCATACTGTGAG 406
 QY 61 GAAAGGGCCAAAGATGACAGAGATTCCTTACGCTGAAAGCCATAGCGGATTTATCT 120
 Db 407 GAAAGGGCCAAAGATGACAGAGATTCCTTACGCTGAAAGCCATAGCGGATTTATCT 466
 QY 121 GGTGATCAAGCAGAGAACTTTTTCATCTGAGTACCTCAGCCTGCTTACAGACA 180
 Db 467 GGTGATCAAGCAGAGAACTTTTTCATCTGAGTACCTCAGCCTGCTTACAGACA 526
 QY 181 ATATGGGCGCTTACCGGACATGAATTAAGATGAAGATGATCAAGTGAATTTTTCATA 240
 Db 527 ATATGGGCGCTTACCGGACATGAATTAAGATGAAGATGATCAAGTGAATTTTTCATA 586
 QY 241 GCCATGAAGCTATCAAACTGAAGCTACAGATATACCTCCCTCCACATTCGCCCT 300
 Db 587 GCTATGAAGCTATCAAACTGAAGCTACAGATATACCTCCCTCCACATTCGCCCT 646
 QY 301 GTCATGAAGACAGCAACAGTGGCTATTTCCAGTGCACAGCAATTTGGTATAGAGGAT 360
 Db 647 GTCATGAAGACAGCAACAGTGGCTATTTCCAGTGCACAGCAATTT----- 691
 QY 361 GCTAGCAGTGCACCACTCAAGCTGTGCTCTGTGCAATGGGCTTCATTCAGTTGTT 420
 Db 692 ----- 691
 QY 421 GGAATGTCTCAACCTTAGTATCTTGTGCTCCAGACAGCAGTGGCTCCCTGGTTAC 480
 Db 692 -----GCAGCTGTGCTCCCTGGTTAC 715
 QY 481 GGGGCTCTCCCTGATACAGCCTGCTGCTGCTTGGGATTCCTGACAGCAATGGCCA 540
 Db 716 GGGGCTCTCCCTGATACAGCCTGCTGCTGCTTGGGATTCCTGACAGCAATGGCCA 775
 QY 541 AAGAGTTCTTCTTACAGAGATCTGCTCAGGGTCACAATTAACAATAAGTTACAGAG 600
 Db 776 AAGAGTTCTTCTTACAGAGATCTGCTCAGGGTCACAATTAACAATAAGTTACAGAG 835
 QY 601 GACACATATCTGAGTGTGCGAGCGCCCTCCAGACAGCAAGTGGGCTGTGCTCAGTA 660
 Db 836 GACACATATCTGAGTGTGCGAGCGCGCCCTCCAGACAGCAAGTGGGCTGTGCTCAGTA 895
 QY 661 TCAAGCTGAAATATACAGGAGTTATTCACAGCAGCAAAAATAATAGTGGACACTTA 720

Db 896 TCAAGCTGAAATATACAGGAGTTATTCACAGCAGCAAAAATAATAGTGGACACTTA 955
 QY 721 AAGAGTTCTTCTTACAGAGATCTGCTCAGGGTCACAATTAACAATAAGTTACAGAG 780
 Db 956 AAGAGTTCTTCTTACAGAGATCTGCTCAGGGTCACAATTAACAATAAGTTACAGAG 1015
 QY 781 TCAATATGAATCTTTCGACATGATCAAGTGAAGAACTACAGCAGCAAAAATAATTC 840
 Db 1016 TCAATATGAATCTTTCGACATGATCAAGTGAAGAACTACAGCAGCAAAAATAATTC 1075
 QY 841 CTAGCTATGACCTTAATGATGTTCCATGTCGTCACCACTGCGCCCTGCTGCT 900
 Db 1076 CTAGCTATGACCTTAATGATGTTCCATGTCGTCACCACTGCGCCCTGCTGCT 1135
 QY 901 CCAAGATATCATCTCTCTTCTTCAAGAGTGTGCTCCGCAAGTGGATGTCCTCAT 960
 Db 1136 CCAAGATATCATCTCTCTTCTTCAAGAGTGTGCTCCGCAAGTGGATGTCCTCAT 1195
 QY 961 AGCTCTTCTGTCGATCAAGGCTGCTGAGAGCGCTGCTCAGAGATGAGAGCAG 1020
 Db 1196 AGCTCTTCTGTCGATCAAGGCTGCTGAGAGCGCTGCTCAGAGATGAGAGCAG 1255
 QY 1021 C---CAGAGAAAGAACTGCTGTCATTTGAGATTAAGAAAGCGGAGAACTTCAGAGCA 1077
 Db 1256 CAATTAAGAAAGAACTGCTGTCATTTGAGATTAAGAAAGCGGAGAACTTCAGAGCT 1315
 QY 1078 GCGAGTGTGAGCTGTGAGAGAGCGCCGCAAGCGCTCTTGGAGCAGCAGCGAAAGCAG 1137
 Db 1316 GCGAGTGTGAGCTGTGAGAGAGCGCCGCAAGCGCTCTTGGAGCAGCAGCGAAAGCAG 1375
 QY 1138 GAGCGGTTGCTCAGCTGAGAGCGCCGCAAGCGCTCTTGGAGCAGCAGCGAAAGCAG 1197
 Db 1376 GAGCGGTTGCTCAGCTGAGAGCGCCGCAAGCGCTCTTGGAGCAGCAGCGAAAGCAG 1435
 QY 1198 CAGAGGCGCAAGCGGCACTGAGAGCGCTGAGAGAGCGCTGAGAGAGCGGAGCTGAG 1257
 Db 1436 CAGAGGCGCAAGCGGCACTGAGAGCGCTGAGAGAGCGCTGAGAGAGCGGAGCTGAG 1495
 QY 1258 CCGCAGCGAGAGAGAGAGAGAGAGAGAGATCGAGAGCGCGGAGCGCGAAACGGGAA 1317
 Db 1496 CCGCAGCGAGAGAGAGAGAGAGAGAGAGATCGAGAGCGCGGAGCGCGAAACGGGAA 1555
 QY 1318 CTGGAAGGCGAGCGCACTTGAATGGGAGCGGAACCGGAGCAGGAATCTGATACAG 1377
 Db 1556 CTGGAAGGCGAGCGCACTTGAATGGGAGCGGAATCGGAATCTGATACAG 1615
 QY 1378 AGGAACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1437
 Db 1616 AGGAACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1675
 QY 1438 GAGTTAGAGCTCTGAATGACAAAGAGATCAGCTAGAGAGAACTTCAGAGATACAG 1497
 Db 1676 GAGTTAGAGCTCTGAATGACAAAGAGATCAGCTAGAGAGAACTTCAGAGATACAG 1735
 QY 1498 TGTGCACTGCGAAGCCAGAGCGAGAAATTTGAGAGCAGGAACAGTCTAGAGAGCTA 1557
 Db 1736 TGTGCACTGCGAAGCCAGAGCGAGAAATTTGAGAGCAGGAACAGTCTAGAGAGCTA 1795
 QY 1558 ATTGCTGAATCAACCACTTACAGCAGAGTTGAGAGATCTCAGCAATGCTTGAAGA 1617
 Db 1796 ATTGCTGAATCAACCACTTACAGCAGAGTTGAGAGATCTCAGCAATGCTTGAAGA 1855
 QY 1618 CTTATTCAGAGAAACAGATATCTCAGTACCAAGTTAAACAGTCCAGCAGAAAGTTG 1677
 Db 1856 CTTATTCAGAGAAACAGATATCTCAGTACCAAGTTAAACAGTCCAGCAGAAAGTTG 1915
 QY 1678 CATAGAGCTGCTCTTACCTTCAAGAGCTTGAAGAGAGAGAGAGAGAGAGAGAGAG 1737
 Db 1916 CATAGAGCTGCTCTTACCTTCAAGAGCTTGAAGAGAGAGAGAGAGAGAGAGAGAG 1975
 QY 1738 CAGCTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1797

Db 1976 CACCTACAGACCAACTGATGAAAGTGAAGAAAGAACTAGATCAAACTACAGAGATT 2035
 Qy 1798 GATGTTTCAACAACCGAGTGAAGAACTGAGAGATACATAGCAAAACAGCACTCCAG 1857
 Db 2036 GATATTTTCAATTAATCAGCTGAGAGAACTAAGAGAAATACCAATTAACCAAACTCCAG 2095
 Qy 1858 AAGCAGAGTCCCTGAGAGCAGCGGAGTGAAGCAG 1893
 Db 2096 AAGCAAAAGTCCATGAGAGGCTGAACGACTGAAGCAG 2131

RESULT 15
 AA234574
 ID AA234574 standard; cDNA; 3231 BP.
 XX AA234574;
 AC
 XX 01-FEB-2000 (first entry)
 DT
 XX Human SH3D1A cDNA clone 9.
 DE
 XX SH3D1A gene; human; Down's syndrome; leukaemia; cancer;
 KM megakaryocytic abnormality; myeloproliferative disorder;
 KM platelet disorder; neural disorder; thrombocytopenia;
 KM haematopoietic disorder; cognitive dysfunction; microcephaly;
 KM lissencephaly; colpocephaly; holoprosencephaly; diagnosis; therapy;
 KM ss.
 XX
 OS Homo sapiens.
 FH
 FT Key Location/Qualifiers
 FT CDS 2..1927
 FT /*tag= a
 XX MO9953062-A2.
 XX 21-OCT-1999.
 XX 16-APR-1999; 99WO-US08371.
 XX 16-APR-1998; 98US-0082007.
 PA (CEDA-) CEDARS SINAI HEAL/TH SYSTEM.
 PI Korenberg JR, Chen X;
 DR WPI; 1999-633829/54.
 DR P-PSDB; AAY32158.
 PT Nucleic acid from the human SH3D1A gene and its products; useful for
 PT the diagnosis and treatment of myeloproliferative disorders and
 PT leukaemia -
 PS
 PS Claim 2; Fig 14; 99P; English.

This is the nucleotide sequence of a non-full-length cDNA (clone 9) corresponding to a novel human SH3 gene, termed the SH3D1A gene, that contributes to the development of platelets and the pathogenesis of leukaemias, both in general and in particular those involving the megakaryocytic lineage. The SH3D1A gene maps to the small candidate region for low platelets on chromosome 21. Sequencing of 5 different sizes of cDNA clone from foetal brain (see AA234570-74) suggests that at least 3 isoforms exist. The invention provides methods for the diagnosis and treatment of megakaryocytic abnormality, myeloproliferative disorder, platelet disorder, acute leukaemia, neural disorders, thrombocytopenia, platelet disorder on chromosome 21, low platelets in deletion for 21, association of gains in chromosome 21 with leukaemias, neural abnormalities, dysfunctions and disorders including brain malformations and corresponding cognitive dysfunctions, microcephaly, lissencephaly, and colpocephaly. Methods are also provided for: suppressing cells unable to regulate themselves; screening for a somatic alteration in the SH3D1A gene; monitoring

CC the progress and adequacy of a treatment; monitoring tumour risk
 CC progress or megakaryocytic abnormality, myeloproliferative disorder,
 CC haematopoietic disorder, platelet disorder or leukaemia; and
 CC treatment of a subject (including a prenatal subject) having
 CC megakaryocytic abnormality, myeloproliferative disorder,
 CC platelet disorder, leukaemia or neural disorder using a
 CC nucleic acid that expresses SH3D1A or its antisense nucleic acid.
 XX
 SO Sequence 3231 BP; 1004 A; 721 C; 712 G; 794 T; 0 other;

Query Match 34.4%; Score 1252.8; DB 20; Length 3231;
 Best Local Similarity 77.2%; Pred. No. 376-289;
 Matches 1652; Conservative 0; Mismatches 272; Indels 216; Gaps 2;

Qy 1506 GGCACCCAGAGGCGAAGAAATTGAGACGACGAACTAGAGCTTAAGATTTGCTGA 1565
 Db 1 GACCAACCCAAAGGCAAGAAATTGAGACGACGAACTTAAGAGTTGGAATTTGCCGA 60
 Qy 1566 AATCACCCTTACAGCAGCAGGTTGAGAACTTCAGCAAAATGCTTGAAGACTTATTC 1625
 Db 61 AATCACCCTTACAGCAGCAGCAGGTTGAGAACTTCAGCAAAATGCTTGAAGACTTATTC 120
 Qy 1626 AGAGAAACAGATTAATCTGATGACGATTAAACAAAGTCCAGCAGAACTTTGCTAGA 1685
 Db 121 AGAAAAACAGTACTCAATGACCAATTAACAAAGTCCAGCAGAACTTTGCTAGA 180
 Qy 1686 CTGCTCTTACCCCTCAAAAGAGCTTGAAGCAAAAGAGTGGCCGGAGCAGCTCCG 1745
 Db 181 TTCCTTGTTCACCTTAAGAGCTTGAAGCAAAAGAGTGGCCGGAGCAGCTTACG 240
 Qy 1746 GAGACAGCTGACAGGTTGAGAGAGACAGGTCAAAGCTGAGAGATTTGTTTT 1805
 Db 241 AGACCAACTGGATTAATGTGAGAAAGAACTAAGTCAAACTACAGAGATTTATTTT 300
 Qy 1806 CAACACCAAGTGAAGAACTGAGAGATTCATGCAAAACGCAATCCAGAAACGAG 1865
 Db 301 CAATTAATCAGCTGAAGAACTTAAGAGAAATACATTAAGCAACATCCAGAAACAAA 360
 Qy 1866 GTCCCTGGAGGCGAGGAGCTGAAGCAAAAGCAGAGAGAGAGGCTTGAGTTAGA 1925
 Db 361 GTCCATGAGGCTGAACGACTGAACAGAAACAGAAACAGAAATCATTAATTTAGA 420
 Qy 1926 GAAACAAAGAGAGACGCTCAGAGACGATTCAGAGAAAGGACAAAGATGCTGAGCA 1985
 Db 421 AAAACAAAGAGAGAGACGCTCAGAGACGATTCAGAGAAAGGACAAAGATGCTGAGCA 480
 Qy 1986 TGTGACAGAGAG--GAGCAGCAGCCCGCGGAAACCCACAGAGAGCAGACTGAA 2042
 Db 481 TGTGACAGAGAGAGCAGCATCAGAGCCAGAAACCTCCACAGAGAGAAATGTAATG 540
 Qy 2043 GAGGGAAGACAGTCTCAGAGAAAGAGGCGGAAAGAGAGCCAAAGCCGGAATGCAAGA 2102
 Db 541 AAGGAGAGAGTCTCAAAAGAGATGCGAGAGAAAGGCAACAGAGAGCAAGAGA 600
 Qy 2103 CAACAGAGTCCGCTTTTCCATCCGATCAGAGACGAGCTTAAGCTGACCAACCGAGCACC 2162
 Db 601 CAACCTGGTGGCTTTTCCATCAACACAGAACGAGCTTAAGCTGACGAGCAGCACC 660
 Qy 2163 CTGCTTCAACAGAGAAAGGCGCTTACCATTTTGCACAGAGAGTGTAAAGTGT 2222
 Db 661 CTGCTTCAACAGAGAAAGGCGCTTACCATTTTGCACAGAGAGATGTAAAGTGT 720
 Qy 2223 ATATTACGAGCGGTGTAACCCCTTTGTAATCCAGAAAGTCAGATGATCCATCCAGGCC 2282
 Db 721 GTATTACCGGCGACCTGTAACCCCTTTGTAATCCAGAAAGTCAGATGATCCATCCAGGCC 780
 Qy 2283 AGAGATATATGTCATGTGTGATGAAGACCAAGCTGAGAGCCAGATGCTTTGAGAGAGA 2342
 Db 781 AGGAGACATATGATGTGTGATGAAGACCAAGCTGAGAGCCAGATGCTTTGAGAGAGA 840
 Qy 2343 GCTGAAGGAGAGAGCGAGTGTTCCTGCAAACTATGACGAGAAAGATTCCAGAAATGA 2402
 Db 841 ATTAAAGGAGAGAGCGAGTGTTCCTGCAAACTATGACGAGAAAGATTCCAGAAATGA 900

```

QY 2403 GGTTCCTCCAGCCAAACAGTGCATCTGACATCTGCCCCCTGCCCCAAACTGAC 2462
Db 901 GGTTCCTCCAGTGAACAGTGCATCTGACATCTGCCCCCTGCCCCAAACTGAC 960
QY 2463 TCTGCGTGAAGACCCCTGCTCTCTTGGCACTGACCTCTTGTGAGCCCTCCACAACCCCA 2522
Db 961 CTTCGCGTGAAGACCCCTGCTCTCTTGGCACTGACCTCTTGTGAGCCCTCCACAACCCCA 1020
QY 2523 CAACGTGGGAGACCTTCAGTTCCACGTGGGCCAGCACTCAAAAGAGAACCCGAAACGGA 2582
Db 1021 TTAACGTGGGCCGACCTTCAGTTCCACGTGGGCCAGCACTCAAAAGAGAACCCGAAACGGA 1080
QY 2583 CAACGTGGGATAGTGGGCCGACCTTCAGCTTCTGACCTGATAGTGGGCCAGTATAG 2642
Db 1081 TTAACGTGGGATAGTGGGCCAGCCAGCCCTCTCTCAGCTTCCAAAGTGGGCCGACGTTAG 1140
QY 2643 GCAGAGATCAGCTTTACCCCAAGCCAGCCACTGCTCTCCCACTCTCCGCTCTGGG 2702
Db 1141 GCAGAGATCAGCTCTTACTCCAGCCAGCCAGCTGCTCTCCGCTCTCTGCTAGG 1200
QY 2703 CCAGGCTGAAAAGTGGAGGGGTAGAACGCAAGCCCTGTATCCTTGGAGAGCCAAAA 2762
Db 1201 CCAGGCTGAAAAGTGGAGGGGTAGAACGCTCAAGCCCTATATCTTGGAGAGCCAAAA 1260
QY 2763 AGACAACCACTTAATTTTAAACAAAAGTACGTCATCACCGTTCTGGAACAGAAAGACAT 2822
Db 1261 AGACAACCACTTAATTTTAAACAAAATATGTATCATCCGTCCTGGAACAGAAAGACAT 1320
QY 2823 GTGTGCTTTGGAGAAATCAAAGTCAAGAGGTTGGTTCCCAAGTCTTAAGTAACT 2882
Db 1321 GTGTGCTTTGGAGAAATCAAAGTCAAGAGGTTGGTTCCCAAGTCTTAAGTAACT 1380
QY 2883 CATTTCAAGGCCCCGTAGAGAAATCCACAGCATCGATCTGCGCCCTACTGAAAGTCTGC 2942
Db 1381 CATTTCAAGGCCCCCATAGAGAAATCTACAGCATGTCTGCTTTCAGAGAGTCTGC 1440
QY 2943 TAGTCTAAAGAGTGGCTTCCCGGCGCCAGCCAGCCGATTTCCCGAGAGAGTTAT 3002
Db 1441 TAGTCTAAAGAGTGGCTCTTCCAGAGCCAGCCGCTGTTTCGAGAGAA----- 1492
QY 3003 TGCCATGTACACATACGAGAGTCTGAGCAAGAGATTTAACCTTTCAGCAAGGGGATGT 3062
Db 1493 ----- 1492
QY 3063 GATTGTGTTACCAAGAAAGATGTGACTGTGACGGGAAGGTGGCGACAAGTCCGG 3122
Db 1493 ----- 1492
QY 3123 AGTCTTCCTCTAATCTATGTAGAGGCTTAAGATTCAGAGGGCTTGAACTGTGGGA 3182
Db 1493 ----- 1492
QY 3183 AACAGGAGATTAGGAAAAAAACCTGAATTTGCCAGTTATGCTTCTCAAGCTGCTAC 3242
Db 1493 -----GAAATTGCCAGTTATTTGCTCATACCGCCAC 1527
QY 3243 TGTCTCCGAACTACCTGCTGCTCTTGGGCACTGATTTGATTCGGAAGAAAGAACCC 3302
Db 1528 CCGCCCCGAGCAGCTCACTCTGCCCCCTGTGAGTGTGATTCGAAAAAGAACCC 1587
QY 3303 AGGTGGATGGTGGGAAAGAGAACTGCAAGCTGAGGGGAAAAAGCCGAGATAGGGGTGTT 3362
Db 1588 AGGTGGATGGTGGGAAAGAGAGCTGCAAGCTGAGGGGAAAAAGCCGAGATAGGGGTGTT 1647
QY 3363 TCCAGCAATTAATGTCAAACTTCAAGCCCGGAAACAGCAAAATCACCCCAACTGAGCT 3422
Db 1648 CCAGCTAATTAATGTCAAACTTCAAGCCCGTGGGAGAGCAAAATCACCTCAACAGAGCC 1707
QY 3423 ACCCAAGACCGCAGTCAAGCAGCAGTGTGCGCAGGTGATCGGATGTACGATTAACCGGC 3482
Db 1708 ACCTAAGTCAACAGCATTAAGCGCAGTGTGCGCAGGTGATGTGAGTGTACGACTACACGC 1767

```

```

QY 3483 CCAGAACGATGACGAAGTACCTTCCAGCAAAAGCCAGATCATCAAGTCTCAACAGGA 3542
Db 1768 GAGAAATGACGATGAGTGGCTTCAACAGGGCCAGATCATCAAGTCTCAACAGGA 1827
QY 3543 GAGCCCGACGTGGTGAAGAGAGAAATGCAAGTGGGCAAGTGGGCTTTCCATCAATTA 3602
Db 1828 GAGCCCTGACTGTGGAAAGGAGAAATGCAATGCAAGTGGGGCTTTCCATCAATTA 1887
QY 3603 TGTAAAGTGCACACAGCATGAGACCCAGCCAGCAATGA 3642
Db 1888 TGTGAAGTGCACACAGCATGAGACCCAGCAATGA 1927

```

Search completed: December 4, 2003, 15:50:26
 Job time : 608.079 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 4, 2003, 15:24:15 ; Search time 4655.38 Seconds
(without alignments)
19013.884 Million cell updates/sec

Title: US-09-674-237a-2

Perfect score: 3642

Sequence: 1 atggtcagcttccaccacaccc.....tgaccaccagcagcaatga 3642

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gse_hum:*
18: em_gse_inv:*
19: em_gse_pin:*
20: em_gse_vrt:*
21: em_gse_fun:*
22: em_gse_mam:*
23: em_gse_mus:*
24: em_gse_pro:*
25: em_gse_rtd:*
26: em_gse_phg:*
27: em_gse_vrt1:*
28: gb_gse1:*
29: gb_gse2:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1496.6	41.1	2079	11	BC013578 Homo sapi
2	1495.	41.0	2126	11	BC020269 Homo sapi
3	809.8	22.2	818	14	CA324450 UI-M-FY0-
4	795.6	21.8	911	13	BQ942708 AGENCOURT

Result No.	Score	Query Match	Length	ID	Description
5	748.8	20.6	2589	11	AK082606 Mus muscu
6	727.2	20.0	732	14	CA750495 UI-M-FY0-
7	724	19.9	725	14	CB248849 UI-M-EXO-
8	722.8	19.8	732	12	BM950476 UI-M-EHOP
9	721.2	19.4	738	12	BM944544 UI-M-EHOP
10	704.8	19.4	729	13	BM704308 UI-M-FOO-
11	701.8	19.3	706	14	CB521237 UI-M-GHO-
12	699.4	19.2	703	14	CB527154 UI-M-FY0-
13	691.2	19.0	713	14	CA329050 UI-M-FY0-
14	684.8	18.8	688	14	CB244555 UI-M-FY0-
15	658.2	18.1	979	13	BQ719508 AGENCOURT
16	656.6	18.0	690	13	BQ179493 UI-M-EWO-
17	654.4	18.0	687	14	CB521525 UI-M-GHO-
18	635	17.4	636	14	CD347867 UI-M-FY0-
19	633.6	17.4	650	10	BB656585 UI-M-FY0-
20	625	17.2	636	14	CB526899 UI-M-FY0-
21	623	17.1	782	13	BQ614534 UI-M-EVO-
22	609.6	16.7	616	29	CC325581 UI-M-EVO-
23	606.8	16.7	610	13	BQ443095 UI-M-EVO-
24	602	16.5	603	9	AI594919 mJ79g10.Y
25	598	16.4	598	14	CA873077 K0922F11-
26	593.2	16.3	606	10	AM910367 UI-M-FY0-
27	592.6	16.3	957	14	CB587690 AGENCOURT
28	583.8	16.0	587	14	CB724039 UI-M-FY0-
29	581.8	16.0	585	14	CB518599 UI-M-FY0-
30	580.4	15.9	582	29	CC183329 XE775.Bay
31	577.4	15.9	579	14	CA879211 K0972F03-
32	572	15.7	572	14	CA895623 B0133D03-
33	571.4	15.7	573	14	CA874698 K0934C04-
34	565	15.5	750	14	CD356862 AGENCOURT
35	563.4	15.5	773	12	BM681943 UI-E-E01-
36	558.8	15.3	639	13	BQ201258 UI-R-DO1-
37	555	15.2	1062	13	BQ898718 AGENCOURT
38	550.2	15.1	551	14	CA463686 K0316G09-
39	550	15.1	751	14	CA448081 UI-H-ED1-
40	547	15.0	906	13	BU519029 AGENCOURT
41	545.4	15.0	547	14	CA565513 K0339A05-
42	539.6	14.8	608	14	CA511580 UI-R-F00-
43	537.8	14.8	724	10	BR373101 601224527
44	534.8	14.7	722	14	CB521233 UI-M-GHO-
45	534.8	14.7	728	14	BY731854 BY731854

ALIGNMENTS

RESULT 1
BC013578
LOCUS
DEFINITION Homo sapiens, similar to interbeta 1 (SH3 domain protein), clone IMAGE:3878242, mRNA.
ACCESSION BC013578
VERSION BC013578.1 GI:15488896
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
1 (bases 1 to 2079)
Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (04-SEP-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK
COMMENT NIH-MGC Project URL: <http://mgc.ncl.nih.gov>
Contact: MGC help desk
Email: cgapbs-remail.nih.gov

Tissue Procurement: DCTD/DP/GenZdar
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305

Web site: <http://www-shgc.stanford.edu>
 Contact: (Dickson, Mark) mcdpaxil.stanford.edu
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ILM, at: <http://image.jlml.gov>
 Series: IRAX Plate: 14 Row: c Column: 5
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA g1: 4504796
 This clone has the following problem: retained intron.

FEATURES

source

1..2079

Location/Qualifiers

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:3878242"

/tissue_type="Lung, large cell carcinoma"

/clone_id="NIH_MGC_68"

/lab_host="DH10B"

/note="Vector: pCMV-SPORT6"

BASE COUNT 703 a 451 c 492 g 433 t

ORIGIN

Query Match 41.1%; Score 1496.6; DB 11; Length 2079;
 Best Local Similarity 86.0%; Pred. No. 0;
 Matches 1671; Conservative 0; Mismatches 269; Indels 3; Gaps 1;

1 ATGGCTCATGTTCCACACCTTTGGGTGATGCTGATGTCGTGGCCATTAATCTGTGAG 60
 Db 132 ATGGCTCATGTTCCACACCTTTGGGTGATGCTGATGTCGTGGCCATTAATCTGTGAG 191
 Qy 61 GAAAGGGCCAGATGACACCAAGTCTTACCTTACCTGAAAGCCGATAGCGGATTTATTA 120
 Db 192 GAAAGAGGAGAGATGATGACACCAAGTCTTACCTTACCTGAAAGCCGATAGCGGATTTATTA 251
 Qy 121 GGTATCAAGGAGAGATTTTTCATCTGGGTGATGCTGATGTCGTGGCCATTAATCTGTGAG 180
 Db 252 GGTATCAAGGAGAGATTTTTCATCTGGGTGATGCTGATGTCGTGGCCATTAATCTGTGAG 311
 Qy 181 ATATGGGGCTAGCGGAGATTAACGATGAAAGATGATCAAGTGAATTTTCCATA 240
 Db 312 ATATGGGGCTAGCGGAGATTAACGATGAAAGATGATCAAGTGAATTTTCCATA 371
 Qy 241 GCCATGAAGCTTATCAACCTGAAGTAAAGATTAACCTCCCTCCCACTTCCCTCCCT 300
 Db 372 GCTATGAAACTTATCAACCTGAAGTAAAGATTAACCTCCCTCCCACTTCCCTCCCT 431
 Qy 301 GTCATGAAAGCAACCAAGTGGTATTTCCAGTGACCAAGCATTTGGTATGAGAGGAT 360
 Db 432 GTCATGAAAGCAACCAAGTGGTATTTCCAGTGACCAAGCATTTGGTATGAGAGGAT 491
 Qy 361 GCTAGCATGCCACCACTCAGAGCTGTGCTCTGTGCAATGGGCTCCATTCAGTTGTT 420
 Db 492 GCCAGCATGCCACCAAGTGGTATTTCCAGTGACCAAGCATTTGGTATGAGAGGAT 551
 Qy 421 GGAATGTTCCACCTTATGATTTCTTCTGCTCCAGAGCAAGTGGCTCCCTGGCTAAC 480
 Db 552 GGAATGTTCCACCTTATGATTTCTTCTGCTCCAGAGCAAGTGGCTCCCTGGCTAAC 611
 Qy 481 GGGGCTCTCCGTCATACAGCTCTGCTGCTGGTGGAGATCCCTGACCAAGTGGCA 540
 Db 612 GGGGCTCTCCGTCATACAGCTCTGCTGCTGGTGGAGATCCCTGACCAAGTGGCA 671
 Qy 541 AAGAGTCTCTCCAGAGATCTGTGCTCAGAGTCAACATTAACACTTAAGTTACAGAG 600
 Db 672 AAGAGTCTCTCCAGAGATCTGTGCTCAGAGTCAACATTAACACTTAAGTTACAGAG 731
 Qy 601 GCAAGATATTCATGATGTCAGAGGCGCTCCAGAGCAAGATGGGCTGTGCTCAGTCA 660
 Db 732 GCAAGATATTCATGATGTCAGAGGCGCTCCAGAGCAAGATGGGCTGTGCTCAGTCA 791
 Qy 661 TCAAGGCTGAATATCAGGCAATTTTCAAGCAAGCAAAACATATGAGTGAACCTTA 720

Db 792 TCAAGATGAATATCAGGCAATTTTCAATGATGATACCAAACTATGAGTGAACCTTA 851
 Qy 721 ACAGGTCCTCCAGCAAGAACTATTTCTATGATCAATCAAGTTTACCCAGGCTCAGTGCT 780
 Db 852 ACAGGTCCTCCAGCAAGAACTATTTCTATGATGATGATGATGATGATGATGATGATGAT 911
 Qy 781 TCAATATGAAATCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840
 Db 912 TCAATATGAAATCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 971
 Qy 841 CTAGCTATGACCTTAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 900
 Db 972 CTAGCTATGACCTTAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1031
 Qy 901 CCAGAAATACATCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 960
 Db 1032 CCAGAAATACATCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1091
 Qy 961 AGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1020
 Db 1092 AGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1151
 Qy 1021 C---CAGAGAAAGAACTGCTGTCATTTGAAGATTAAGAAAGCGGAGAACTTGCAGCGA 1077
 Db 1152 CAATTAGAAAGAAAGAACTGTCATTTGAAGATTAAGAAAGCGGAGAACTTGAAGCT 1211
 Qy 1078 GGCAGTGTGAGCTGAGAAAGCCGCGCAAGCTCTTGAAGACAGCCGCAAGAGCAG 1137
 Db 1212 GGCAGTGTGAGCTGAGAAAGCCGCGCAAGCTCTTGAAGACAGCCGCAAGAGCAG 1271
 Qy 1138 GAGCGTGTGAGCTGAGAAAGCCGCGCAAGCTCTTGAAGACAGCCGCAAGAGCAG 1197
 Db 1272 GAGCGTGTGAGCTGAGAAAGCCGCGCAAGCTCTTGAAGACAGCCGCAAGAGCAG 1331
 Qy 1198 CAGGAGGCAAGCGGAGCTGAGCTGAGAAAGCTGAGAAAGCTGAGAAAGCTGAGAAAGCTGAGAA 1257
 Db 1332 CAGGAGGCAAGCGGAGCTGAGCTGAGAAAGCTGAGAAAGCTGAGAAAGCTGAGAAAGCTGAGAA 1391
 Qy 1258 CCGCAGCGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1317
 Db 1392 CCGCAGCGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1451
 Qy 1318 CTGGAAGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1377
 Db 1452 CTGGAAGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1511
 Qy 1378 AGGAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1437
 Db 1512 AGGAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1571
 Qy 1438 GAGTTAGAGCTGTAATGACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1497
 Db 1572 GAGTTAGAGCTGTAATGACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1631
 Qy 1498 TGTGCACTGCAACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1557
 Db 1632 TGTGCACTGCAACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1691
 Qy 1558 ATTGCTGAATTCACCCACTTAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1617
 Db 1692 ATTGCTGAATTCACCCACTTAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1751
 Qy 1618 CTATATTCAGAGAAACAGATATCTAGTGACCAAGTTAAACAAAGTCCAGCAAGAAAGTTTG 1677
 Db 1752 CTATATTCAGAGAAACAGATATCTAGTGACCAAGTTAAACAAAGTCCAGCAAGAAAGTTTG 1811
 Qy 1678 CATAGAGACTGCTTCTTAACTTAAAGAGCTTGAAGCAAGAGAGAGAGAGAGAGAGAGAGAGAG 1737
 Db 1812 CATAGAGACTGCTTCTTAACTTAAAGAGCTTGAAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1871
 Qy 1738 CAGCTCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1797

Db 1872 CACCTACGAGACCACTGATGAGTGAAGAGAGAGAGAGAGAGAGAT 1931
 QY 1798 GATGTTTCAACACACAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1857
 Db 1932 GATATTTTCAATATATAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAG 1991
 QY 1858 AAGCAGAGGTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1917
 Db 1992 AAGCAAAAGTCCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2051
 QY 1918 GAGTTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1940
 Db 2052 GAATTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2074

RESULT 2
 BC020269 2126 bp mRNA linear HTC 19-DEC-2001
 LOCUS BC020269 Homo sapiens, clone IMAGE:4899011, mRNA.
 DEFINITION BC020269
 ACCESSION BC020269.1 GI:17939664
 VERSION
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 2126)
 Strausberg, R.
 Direct Submission
 Submitted (19-DEC-2001) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 NIH-MGC Project URL: http://mgc.nci.nih.gov
 CONTACT: MGC help desk
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Rubin Laboratory
 DNA Sequencing by: Genome Sequence Centre,
 BC Cancer Agency, Vancouver, BC, Canada
 info@cgsc.bc.ca
 Steven Jones, Jennifer Amano, Ian Bosdet, Yaron Butterfield,
 Susana Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
 Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
 Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
 Ness, Pawan Pandoh, Anna-Lisa Prabhu, Parvaneh Saedi, Jacqueline
 Schein, Duane Smalhus, Michael Smith, Lorraine Spence, Jeff Stott,
 Michael Thorne, Miranda Tsai, Natasja van den Bosch, Jill Vardy,
 George Yang, Scott Zuyderduyn, Marco Marra.

REMARK
 COMMENT
 Cloning distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
 Series: IRAL Plate: 40 Row: n Column: 1
 This clone was selected for full length sequencing because it
 passed the following selection criteria: matched mRNA gi: 4504796
 This clone has the following problem: no cloning site /
 microdeletion.
 Location/Qualifiers
 1. 2126
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:4899011"
 /issue_type="Pancreas", epithelioid carcinoma"
 /clone_lib="NIH-MGC 42"
 /lab_host="DH10B-R"
 /note="Vector: pOTB7"

BASE COUNT 712 a 462 c 517 g 435 t
 ORIGIN
 Query Match 41.0%; Score 1495; DB 11; Length 2126;
 Best Local Similarity 85.9%; Pred. No. 0;

Matches 1670; Conservative 0; Mismatches 270; Indels 3; Gaps 1;
 QY 1 AAGGCTCAAGTTTCCACACCTTTCGGTGTACCTGAGAGAGAGAGAGAGAGAG 60
 Db 182 AAGGCTCAAGTTTCCACACCTTTCGGTGTACCTGAGAGAGAGAGAGAGAGAG 241
 QY 61 GAAAGGCGCAAGATGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
 Db 242 GAAAGGCGCAAGATGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 301
 QY 121 GGTATCAAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
 Db 302 GGTATCAAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 361
 QY 181 ATATGGGCGCTGCGGAGATGATGATGATGATGATGATGATGATGATGATGAT 240
 Db 362 ATATGGGCGCTGCGGAGATGATGATGATGATGATGATGATGATGATGATGAT 421
 QY 241 GCGATGAGCTTATCAAACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
 Db 422 GCGATGAGCTTATCAAACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 481
 QY 301 GTGATGAAACAGCAACAGTGGCTATTTCCAGTGCACAGCAATTTGATAGAGAGAT 360
 Db 482 GTGATGAAACAGCAACAGTGGCTATTTCCAGTGCACAGCAATTTGATAGAGAGAT 541
 QY 361 GGTAGATGACCACTACAGCTGTGTCTCTGTGCGCAATGGGCTCCATTCAGTTGTT 420
 Db 542 GGTAGATGACCACTACAGCTGTGTCTCTGTGCGCAATGGGCTCCATTCAGTTGTT 601
 QY 421 GGAATGTCTCCACCTTATGATCTTCTGTCCTGAGAGAGAGAGAGAGAGAGAG 480
 Db 602 GGAATGTCTCCACCTTATGATCTTCTGTCCTGAGAGAGAGAGAGAGAGAGAG 661
 QY 481 GGGGCTCTCTCCGCTATACAGCTGTGTCTGCTGCTGCTGCTGCTGCTGCTGCT 540
 Db 662 GGGGCTCTCTCCGCTATACAGCTGTGTCTGCTGCTGCTGCTGCTGCTGCTGCT 721
 QY 541 AAGAGTTCTCTTTCAGAGAGATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 600
 Db 722 AAGAGTTCTCTTTCAGAGAGATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 781
 QY 601 GCACATCATTCGATGTCGCGCAGCGCCCTCCAGAGAGAGAGAGAGAGAGAGAG 660
 Db 782 GCACATCATTCGATGTCGCGCAGCGCCCTCCAGAGAGAGAGAGAGAGAGAGAG 841
 QY 661 TCAGAGCTGAAATACAGGAGATTTTCAACAGCCAGCAAACTATAGAGAGAGAG 720
 Db 842 TCAGAGCTGAAATACAGGAGATTTTCAACAGCCAGCAAACTATAGAGAGAGAG 901
 QY 721 AAGGTCCTCCAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
 Db 902 AAGGTCCTCCAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 961
 QY 781 TCAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840
 Db 962 TCAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1021
 QY 841 CTAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
 Db 1022 CTAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1081
 QY 901 CAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960
 Db 1082 CAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1141
 QY 961 AGCTGTCCTGTCGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020
 Db 1142 AGCTGTCCTGTCGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1201
 QY 1021 C---CAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1077
 Db 1202 CAATTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1261

OY	1078	GGCAGTGTGGAGCTTGAGAAAGGCGCCGCAAGGCGCTTGTGAGCAGCAGCGCAAAAGACG	1137
Db	1262	GGCAACCTGGAACTGGAGAAACGAAGGCAAGCTCTCTTGAAACAGCAGCGCAAGAGCAG	1321
OY	1138	GAGCGGTGGCTCAGCTGGAAGCGCGCCGAGCAGAGAGAGAAAGACGGAGCGCCAGAG	1197
Db	1322	GAGCGCTGGCCAGCTGGAAGCGGCGGAGCAGAGAGAGAAAGAGAGAGCTGAGCGCCAGAG	1381
OY	1198	CAGAGAGCGCAACGGCAGCTGGAAGCAAGCTTGAGAAAGACACGGAGAGTGGAG	1257
OY	1258	CGCGCAGCGAGAGAGAGAGAGAGAGATCGAGAGCGCGCAGAGCGCGCAAAACGGGAA	1317
Db	1442	CGCGAGAGAGAGAGAGAGAGAGAGAAATTGAGAGCGAGAGCTGCAAAACGGGAA	1501
OY	1318	CTGGAAAGGACGCGCAACTTGAATGGAAACGGAAACGGAGACAGAACTCTGATACG	1377
Db	1502	CTTGAAGGCAACGCAACTTGAATGGGAAACGGAACTGCAAGCAAGACTCTAATACAA	1561
OY	1378	AGGAACAAGAGAGAGAGAGGCAACCGTGTCTTGAAAGCAAGAGAGAACTCTGAGATT	1437
Db	1562	AGAAACAAAGAAACAAGAGACATGATTTGACTGAAAGCAAAAGAAACATTTGGAAATT	1621
OY	1438	GAGTTAGAAGCTCTGAATGACAAAAGACATCAGTAGAAGGAAACCTTCAGATTCAG	1497
Db	1622	GAATTTGAAGCTCTTAATGATTAAGAAAGCATCACTGAAAGGAAACCTTCAGATTCAG	1681
OY	1498	TGTGACTGGCAACCCAGAGGCAAGAAATTGAGACACGAACAACTTAGAGACTTGAAG	1557
Db	1682	TGTCGATTTGACCAACCCAAAGGCAAGAAATTGAGACACGAACAACTTAGAGAGTTGAG	1741
OY	1558	ATTGCGTAATCAACCACTTACAGAGAGTTGCAAGATCTCAGCAATGCTTGGAGAG	1617
Db	1742	ATTGCGAAATCAACCATCTACAGCAATTAACGAATCTCAGCAATGCTTGGAGAG	1801
OY	1618	CTTATTTCAAGAGAAACAGATTACTCAGTACCCAGTTTAAACAAAGTCACAGAACAGTTG	1677
Db	1802	CTTATTTCAAGAGAAACAGATTACTCAATGACCAATTAAACAAAGTTCAACAGAACAGTTG	1861
OY	1678	CATAGAGACTCGCTTCTTAACCTCAAAAGAGCTTTGAAACCAAGAGCTGGCCGGCAG	1737
Db	1862	CACAGAGATTCACCTTGTTACACTTAAAGAGCTTTGAAAGCAAAAGAACTGCTGGCAG	1921
OY	1738	CAGCTCCGGAGCAGCTGGAAGAGGTGAGAGAGACAGGTCAACCTCAGAGAGATT	1797
Db	1922	CACCTACAGACCAATGATGAAATGGAAGAACTGAGTCAAACTACAGAGAGATT	1981
OY	1798	GATGTTTTCAACCAACAGCTGGAAGAACTGAGAGATACATAGCAACAGCAACTCCAG	1857
Db	1982	GATATTTTCAATTAATCACTGAGAGAACTTAAGAGAAATACCAATTAACCAACTCCAG	2041
OY	1858	AAAGCAGAGTCCCTGGAGGCAAGCGGCACTGAAACAGAAAGACAGAGAGAGAGCTGT	1917
Db	2042	AAAGCAAAAGTCCATGAGAGCTGGAACGACTGAAACAGAAAGAAACAGAAAGATCATTA	2101
OY	1918	GAGTTAGAGAGCAAAAGAGAGA	1940
Db	2102	GAATTTGAAAAAAGAAAAA	2124
RESULT 3			
CA324450			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			

REFERENCE	1 (bases 1 to 818)
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/ .
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL	Unpublished
COMMENT	Contact: Robert Strausberg, Ph.D. Email: robert@xena1.nih.gov

Tissue Procurement: Dr. Jim Lin, University of Iowa
 CDNA library preparation: Dr. M. Bento Soares, University of Iowa
 CDNA library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/BLNI at: <http://image.llnl.gov>
 Dr. Jim Lin, University of Iowa

This clone was contributed by the Brain Molecular Anatomy Project (BMAP).
The following repetitive elements were found in this cDNA sequence: 338-455, >(GGA)nSimple_repeat
Seq primer: pYX-5.
Location[chr]:1:1496

FEATURES

Source

Location/Qualifiers
1. .818

```

/mol_type="mRNA"

```

```
/Bcra11n="C5/BL/6"
/db xref="taxon:10090"
```

```
/clone="IMAGE: 6820974
```

```
/dev_stage="embryo 13.5,14
```

```
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH BMAP FY0"
```

```
/note="Organ: Brain; Vector: pYX-Asc;
```

Bonaldo, Lennon and Soares, *Genome Research*, 6:791-806,

1996. First strand cDNA synthesis was primed with 0190-dT

primer containing a Not I site. Double strand cDNA was

with *EcoR* I adaptor, digested with *NotI* and then cloned

directionally into pYX-Abs vector. The library tag sequence located between the Not I site and the polyA tail

is AGCAGACAG. This library was created for the University of

Developing Mouse Nervous System', supported by National

Institute of Mental Health (NIMH), Bethesda, Md., U.S.A.
Program Coordinator "

a	179	c	270	g	127	t	4	others
---	-----	---	-----	---	-----	---	---	--------

22.26; SCORE 809.8; DB 14; Length 818;
99.38; Pred. NO. 2e-171;

conservative 0; Mismatches 6; Indels 0; Gaps 0

TTTATCCTAGCTATGCACCTAATTGATGTTGCCATGTTCTGTGTCAGCCACTGCCG 888

TTTTATTCCTAAGCTATGCAACCTTAATTGATGTTGACCAATGTCATGTCAGCCACATGCGCG 61

.....

[illegible]

CTGGCTCCAGATACATCCCTCTCTCTCTCAGAGATCCCTCCGCACTGG 121

GTCATAAGCTCTTCTGTGGATCAGAGGCTGCCGTGAGGAGCCGTCGTACAGAG 100

GTCTATAGCTCTTCTTCTGTGGATCAGAGGCTGCCCTGAGGAGCCGTCGTCAGAG 181

XXXXXXXXXXXXXXXXXXXXX

[illegible]

CCGAGGCGAGTGTGGAGCTGGAGAGAGCCGCCCAAGCGCTCTTGGAGGAGCAGCGC 112

CGAGGCACTGTGAGCTGGAGAGCGCCGCCAAGCGTCTTGAGCAGCAGCGC 301

CAGGAGCGGTTGGCTCAGCTGGAGCGCGCCGAGCAGAGAGAGGAAGAAGAGCGGAG 118

Db 302 AAAAGCAGAGACCGGTTGGCTTCAGCTGGAGCCCGCCGAGCAGAGAGAGAAAAGCGCGAG 361

QY 1189 CGCCAGAGCAGAGAGGCCCAAGCGGCACTTGAGCTTGAGAAAGCAGCTTGAGAGCGG 1248

Db 362 CGCCAGAGCAGAGAGCGCAAGCGGCACTTGAGCTTGAGAAAGCAGCTGAGAGAGCGG 421

QY 1249 GAGCTGAGCGCGCAGCGAGGAGGAGGAGAGAGAGATCGAAGCGCCGAGGCCCCA 1308

Db 422 GAGCTGAGCGCGCAGCGAGGAGGAGAGAGAGATCGAAGCGCCGAGGCCCCA 481

QY 1309 AAACCGGAACTGGAAAGGCAAGCGCAACTTGAATGAGAACGAAACCGAGACAGGAACTC 1368

Db 482 AAACCGGAACTGGAAAGGCAAGCGCAACTTGAATGAGAACGAAACCGAGACAGGAACTC 541

QY 1369 CTGAATCAGAGGACAGAGAGCAGAGAGGCACTCGTGCTCTGAAAGCAAGAGGAACT 1428

Db 542 CTGAATCAGAGGACAGAGAGCAGAGAGGCACTCGTGCTCTGAAAGCAAGAGGAACT 601

QY 1429 CTGAGTTTGAGTTTGAAGGCTCTGAATGACAAAAGGCAATCGCTGAAAGGAAACTTTCAG 1488

Db 602 CTGAGTTTGAGTTTGAAGGCTCTGAATGACAAAAGGCAATCGCTGAAAGGAAACTTTCAG 661

QY 1489 GATATCAGGTGTGCACTGGCAACCCAGAGGCAAGAAATTGAGAGCAGCAACAAGTCTAGA 1548

Db 662 GATATCAGGTGTGCACTGGCAACCCAGAGGCAAGAAATTGAGAGCAGCAACAAGTCTAGA 721

QY 1549 GAGCTTAAGAAATTGCTGAATATCACCACTTACAGCAGCAGTTGCGAGGAATTTCCAGCAATG 1608

Db 722 GAGCTTAAGAAATTGCTGAATATCACCACTTACAGCAGCAGTTGCGAGGAATTTCCAGCAATG 781

QY 1609 CTTGGAAGACTTATTCAGAGAGAAAGAGATCACTCAAGTG 1645

Db 782 CTTGGAAGACTTATTCAGAGAGAAAGAGATCACTCAAGTG 818

RESULT 4
 LOCUS BQ942708
 DEFINITION BQ942708 911 bp mRNA linear EST 21-Aug-2002
 AGENCOURT 8764676 NIH_MGC_129 Mus musculus CDNA clone IMAGE:6314690
 5', mRNA sequence.
 BQ942708
 BQ942708.1 GI:22358186
 EST.
 Mus musculus (house mouse)
 SOURCE
 ORGANISM
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
 1 (baes 1 to 911)
 NIH-MGC <http://mgc.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 unpublished
 COMMENT
 Contact: Robert Strusberg, Ph.D.
 Email: cgasbs-remail.nih.gov
 Tissue Procurement: Susan L. Sullivan, PhD.
 cDNA Library Preparation: Reegen, Invitrogen Corp
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNLN at:
<http://image.llnl.gov>
 Plate: L14M13743 row: n column: 03
 High quality sequence start: 10
 High quality sequence stop: 564.
 Location/Qualifiers
 1..911

```

/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="Paxon:10090"
/clone="IMAGE:6314690"
/lab_host="DH10B (phage-resistant)"
/clone_1lb="NH MGC 129"
/note="Organ: olfactory epithelium; Vector

```

BASE COUNT	ORIGIN
271 a	210 c 280 g 150 t

Query Match	21.8%	Score 795.6	DB 13	Length 911
Best Local Similarity	98.0%	Pred. No. 3,46	168	
Matches	890	Conservative	0	Mismatches 9
				Indels 9
				Gaps 8
QY 678	GCAGTATTCGACGCCACGACAAACTATGAGTGAACACTTAA	CAGGTCCCGCAGCAG	737	
DB 1	GCAGTATTCGACGCCACGACAAACTATGAGTGAACACTTAA	CAGGTCCCGCAGCAG	60	
QY 738	AACGATCTCATGCGAATCAAGTTTACCCCGAGGCTCAGCTGGCTTCAATATGGAATCTTTC		797	
DB 61	AACGATCTCATGCGAATCAAGTTTACCCCGAGGCTCAGCTGGCTTCAATATGGAATCTTTC		120	
QY 798	TGACATTGATCAAGATGGAATACTCACTGACAGAGATTTATCCTAGCTATGACCTTAAT		857	
DB 121	TGACATTGATCAAGATGGAATACTCACTGACAGAGAGATTTATCCTAGCTATGACCTTAAT		179	
QY 858	TGATGTTGCCATGTCGTGTGACGCCACCTGCCCGCCGTCCTGCTTCAGAAATACATCCTCC		917	
DB 180	TGATGTTGCCATGTCGTGTGACGCCACCTGCCCGCCGTCCTGCTTCAGAAATACATCCTCC		239	
QY 918	TTCCCTCAGAGAGATTGCTCCGACAGTGGGATGTCGGTATTAAGCTCTTCTGTGGA		977	
DB 240	TTCCCTCAGAGAGATTGCTCCGACAGTGGGATGTCGGTATTAAGCTCTTCTGTGGA		299	
QY 978	TCAGAGGCTGCTCTGAGAGGCTGTCGTGAGAGATGAGCAGCAGCCAGAGAGAAACTGCC		1037	
DB 300	TCAGAGGCTGCTCTGAGAGGCTGTCGTGAGAGATGAGCAGCAGCCAGAGAGAAACTGCC		359	
QY 1038	TGTCGACATTTGAAGATTAAGAGGCGGAGAACTTCGACGAGGAGAGTGTGAGCTGAGAA		1097	
DB 360	TGTCGACATTTGAAGATTAAGAGGCGGAGAACTTCGACGAGGAGAGTGTGAGCTGAGAA		419	
QY 1098	GCGCGCCGCAAGCGCTCTTGAGCAGCAGCCGCAAGAGCAGAGCCGTTTGCTCAGCTGA		1157	
DB 420	GCGCGCCGCAAGCGCTCTTGAGCAGCAGCCGCAAGAGCAGAGCCGTTTGCTCAGCTGA		479	
QY 1158	GCGCGCCGACAGAGAGAGAAAGCGGAGCGCCAGAGACAGAGGCTCAAAGCGCAGCT		1217	
DB 480	GCGCGCCGACAGAGAGAGAAAGCGGAGCGCCAGAGACAGAGGCTCAAAGCGCAGCT		539	
QY 1218	GGAAGCTGAGAGACAGCTGTGAGAGAGCAGCGGAGCTGAGACCGGACGAGAGAGAG		1277	
DB 540	GGAAGCTGAGAGACAGCTGTGAGAGAGCAGCGGAGCTGAGACCGGACGAGAGAGAG		599	
QY 1278	GAGGAGAGAGATTCAGAGGCGCGAGCGCCGCAAAACGGGAATCGGAAGGCGACCACT		1337	
DB 600	GAGGAGAGAGATTCAGAGGCGCGAGCGCCGCAAAACGGGAATCGGAAGGCGACCACT		659	
QY 1338	TGAATGGGAAC-GSAAACGGAGACAGGAATCCTGAAATCAGAGAAACAAGGACAGGAG		1396	
DB 660	TGAATGGGAACGGGAACGGAGACAGGAATCCTGAAATCAGAGAAACAAGGACAGGAG		719	
QY 1397	GC--ACCGGTGCTCTGAGGCAAGAGAGAGACTTGTGAGTTTGAATTAGAACTTGAA		1454	
DB 720	GCCACCGGTGCTCTGAGGCAAGAGAGAGACTTGTGAGTTTGAATTAGAACTTGAA		779	
QY 1455	TGACAAATAACATTCAGTAAAGAAATCTTCAGAGATATCAGGT-GTCCATCTGCAACCC		1513	
DB 780	TGACAAATAACATTCAGTAAAGAAATCTTCAGAGATATCAGGTGTCCATCTGCAACCC		839	
QY 1514	AGA-GGCAAGAAATTGAGACAGC-ACAAGTCTAG-GAGCTTAAGATATGC-TGAATC		1569	
DB 840	AGAGGGCAAGAAATTGAGACAGCACAAGTCTTAAGAGGCTTAATAATTGCTTGAATC		899	
QY 1570	ACCCACTT 1577			

Db 900 ACCACTT 907

RESULT 5
AK082606
LOCUS
DEFINITION
AK082606 2589 bp mRNA linear HTC 05-DEC-2002
Mus musculus 0 day neonate cerebellum cDNA, RIKEN full-length
enriched library, clone: C230068L04 product: intersectin (SH3 domain
protein 1A), full insert sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
AK082606
AK082606.1 GI:26349812
HTC, CAP trapper.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
99279253
10349636
Carninci, P., and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
20499374
11042159
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
20530913
11076861
Shibata, K., Itoh, M., Aizawa, K., Nagao, S., Sasaki, N., Carninci, P.,
Kono, H., Akiyama, J., Nishi, K., Kitsu, T., Tashiro, H., Itoh, M.,
Sun, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M.,
Yonekura, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system-384 format
sequencing pipeline with 384 multichannel sequencer
Genome Res. 10 (11), 1757-1771 (2000)

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
20530913
11076861
Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,
Aizawa, T., Hara, A., Fukunishi, Y., Kono, H., Adachi, D., Fukuda, S.,
Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamada, I.,
Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R.,
Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T.,
Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H.,
Kuehl, P., Lewis, S., Matsumoto, Y., Nikaido, I., Pesole, G.,
Quackenbush, J., Schmitt, L. M., Staudt, F., Suzuki, R., Tomita, M.,
Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H.,
Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N.,
Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C.,
Flatcher, C., Fujita, M., Gariboldi, M., Guinacich, S., Hill, D.,
Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P.,
Marchionni, L., Mashima, J., Mazzei, J., Nombarelli, P., Nordone, P.,
Rising, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,
Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suuki, H.,
Toyo-oka, K., Wang, K. H., Weitz, C., Whitaker, C., Wilming, L.,
Wyehaw-Borje, A., Yoshida, K., Hasegawa, Y., Kawai, H., Kohlschki, S.,
and Hayashizaki, Y.
Functional annotation of a full-length mouse cDNA collection
Nature 409 (6821), 685-690 (2001)

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
11217851
The PANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
6 (bases 1 to 2589)
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T.,
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Kono, H., Kouda, M.,
Koya, S., Kurihara, C., Matsumoto, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohnato, N.,
Okazaki, Y., Saito, K., Satoh, H., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
Muramatsu, M., and Hayashizaki, Y.
Direct Submission
Submitted (15-APR-2002) Yoshida Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gsr.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)
CDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.go.jp/
URL: http://fantom.gsc.riken.go.jp/
URL: http://location.gsc.riken.go.jp/
location/Qualifiers

FEATURES
source
1..2589
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="PANTOM:DB:C230068L04"
/db_xref="taxon:10090"
/clone="C230068L04"
/tissue_type="cerebellum"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="0 day neonate"
/dev_stage="0 day neonate"
/note="unnamed protein product; intersectin (SH3 domain
protein 1A) (MGDI|GI:1338069, GB|NM_010587, evidence:
BLASTN 99%, match=2267)
putative"
/codon_start=2
/protein_id="BAC38546.1"
/db_xref="GI:26349813"
/translation="PFRGPRKSTSIDTPTSPASLAKRVASPAKPAIAGEEFIM
VTYSSHQGLTFOGQIVYVTKKDDMMGTGDKSGVPSNVYRLKDSGSGTAGC
TGSIGKRPETAYASAAATGPELTARPOLILIRKNGSGWEGELQARGKROIG
WFPANVYKLSPGSKITPTLPETAPAVCAVGVIGYDTAQNDBELASKGIIN
LNKEDPVMWGEVSGVGLPSPNVKLTMDPQWQCSLDLMDLPTPERKQGYI
HELIVTEENVNDQLVTEIFPXLTESELLTEKEVAMIFVNMELTMCNKLKAR
VRKSGEKKPKWIKGILSAQHPQYRFGSCQNGALLIQKDEAPDFEYK
RLANDPRCKMPKSSFLIKMQRVTRPLIIXNLENTPEHPHSHKALEAEEL
CSQYNEVREKNSDRLEMTQAHYCGSLSEQVPSNVTCIGPRKILSGKLYKAS
NKEVYGLFLLITITTPRSGSGTDYKPSKSNQYKTKPTPLNRYLVLPD
PSGEPYFHSIDRVYTLAESINERTAWQKAKASELYETERKREKAVYVRSD
RATIGLWVNVGELIKPCRSKGNPNCEVYMGQCHITKTIQTLNPKMSNCQ
PFIDLEOVYCTVFERDQLSADDFGRTEIRVADIKDQSGKPVTKILLHEVPT
GEIVRDLQFDEP"

CDS
1..2269
/note="unnamed protein product; intersectin (SH3 domain
protein 1A) (MGDI|GI:1338069, GB|NM_010587, evidence:
BLASTN 99%, match=2267)
putative"
/codon_start=2
/protein_id="BAC38546.1"
/db_xref="GI:26349813"
/translation="PFRGPRKSTSIDTPTSPASLAKRVASPAKPAIAGEEFIM
VTYSSHQGLTFOGQIVYVTKKDDMMGTGDKSGVPSNVYRLKDSGSGTAGC
TGSIGKRPETAYASAAATGPELTARPOLILIRKNGSGWEGELQARGKROIG
WFPANVYKLSPGSKITPTLPETAPAVCAVGVIGYDTAQNDBELASKGIIN
LNKEDPVMWGEVSGVGLPSPNVKLTMDPQWQCSLDLMDLPTPERKQGYI
HELIVTEENVNDQLVTEIFPXLTESELLTEKEVAMIFVNMELTMCNKLKAR
VRKSGEKKPKWIKGILSAQHPQYRFGSCQNGALLIQKDEAPDFEYK
RLANDPRCKMPKSSFLIKMQRVTRPLIIXNLENTPEHPHSHKALEAEEL
CSQYNEVREKNSDRLEMTQAHYCGSLSEQVPSNVTCIGPRKILSGKLYKAS
NKEVYGLFLLITITTPRSGSGTDYKPSKSNQYKTKPTPLNRYLVLPD
PSGEPYFHSIDRVYTLAESINERTAWQKAKASELYETERKREKAVYVRSD
RATIGLWVNVGELIKPCRSKGNPNCEVYMGQCHITKTIQTLNPKMSNCQ
PFIDLEOVYCTVFERDQLSADDFGRTEIRVADIKDQSGKPVTKILLHEVPT
GEIVRDLQFDEP"

BASE COUNT
693 a 662 c 688 g 546 t
ORIGIN

Query Match 20.6%; Score 748.8; DB 11; Length 2589;
Best Local Similarity 99.7%; Pred. No. 1.8e-157;
Matches 750; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

DB 2890 GGCGCCGTAAGAAATCCACAGCATCGATACGCGCCCTAGTGAAGTCTGCTAGTCTA 2949
14 GGCGCCGTAAGAAATCCACAGCATCGATACGCGCCCTAGTGAAGTCTGCTAGTCTA 73

```

QY 2950 AAGAGAGTGGCTTCCCGGCGGCAAGCCAGCCATTCCCGAGAGAGATTATTGCCATG 3009
DB 74 AAGAGAGTGGCTTCCCGGCGGCAAGCCAGCCATTCCCGAGAGAGATTATTGCCATG 133
QY 3010 TACACATAGAGAGTCTGAGCAGAGAGATTAACTTTTCAAGCAGAGAGATGATTG 3069
DB 134 TACACATAGAGAGTCTGAGCAGAGAGATTAACTTTTCAAGCAGAGAGATGATTG 193
QY 3070 GTTACCAAGAAAGATGATGCTGTGGAAGGGAACGGTGGGCGACAAGTCCGAGCTTC 3129
DB 194 GTTACCAAGAAAGATGATGCTGTGGAAGGGAACGGGCGACAAGTCCGAGCTTC 253
QY 3130 CCTTCTACTATGATGAGGCTTAAAGATTCAAGAGGCTCTGGAATCTCTGGGAAAGAGG 3189
DB 254 CCTTCTACTATGATGAGGCTTAAAGATTCAAGAGGCTCTGGAATCTCTGGGAAAGAGG 313
QY 3190 AGTTTGGAAAAAACCCTGAATTTGCCAGGTTATTGCTTCTTCACTGCTGCTACTGTCCT 3249
DB 314 AGTTTGGAAAAAACCCTGAATTTGCCAGGTTATTGCTTCTTCACTGCTGCTACTGTCCT 373
QY 3250 GAACAACTACCTGCTCTGCGGAGCTGATCTGATCTGATCCGAAAAAGAACCCAGGTGGA 3309
DB 374 GAACAACTACCTGCTCTGCGGAGCTGATCTGATCTGATCCGAAAAAGAACCCAGGTGGA 433
QY 3310 TGTGTGGAAAGAGAGACTGCAAGCTCGAGGGAAAAAGCGCAGATAGGCTGTTCCAGCA 3369
DB 434 TGTGTGGAAAGAGAGACTGCAAGCTCGAGGGAAAAAGCGCAGATAGGCTGTTCCAGCA 493
QY 3370 AATTATGTCAAACTTCTTAAGCCCGGAAACAGCAAAATCAACCCCACTGAGCTACCCAG 3429
DB 494 AATTATGTCAAACTTCTTAAGCCCGGAAACAGCAAAATCAACCCCACTGAGCTACCCAG 553
QY 3430 ACCGAGTGCAGCAGCAGTGTGCGCAGGTGATGCGGATGATGATGATGATGATGATGATG 3489
DB 554 ACCGAGTGCAGCAGCAGTGTGCGCAGGTGATGCGGATGATGATGATGATGATGATGATG 613
QY 3490 GATGACGAATAGCTTCTGAGCAAGGCGCAGATCATCAAGCTCTTCAAGAGAGAGAGCCG 3549
DB 614 GATGACGAATAGCTTCTGAGCAAGGCGCAGATCATCAAGCTCTTCAAGAGAGAGAGCCG 673
QY 3550 GACTGTGGAAGAGAGAGTCAAGTGTGCGCAAGTGGGCTCTTCCATTCATTAATGTAAG 3609
DB 674 GACTGTGGAAGAGAGAGTCAAGTGTGCGCAAGTGGGCTCTTCCATTCATTAATGTAAG 733
QY 3610 CTGACCAAGACATGAGACCCCGCAGCCAGCATG 3641
DB 734 CTGACCAAGACATGAGACCCCGCAGCCAGCATG 765

RESULT 6
CA750495 732 bp mRNA linear EST 27-NOV-2002
DEFINITION UI-M-FYO-cde-p-10-0-UI.r1 NIH-BMAP_FYO Mus musculus cDNA clone
ACCESSION CA750495
IMAGE: 6832091 5', mRNA sequence.
VERSION CA750495.1 GI:25575732
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 732)
JOURNAL NIH-MGC http://mgi.nci.nih.gov/
COMMENT National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgaabs-remail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
CDNA Library preparation: Dr. M. Bento Soares, University of Iowa
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:

```

```

http://image.lnl.gov
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)
The following repetitive elements were found in this cDNA
sequence: 637-722, >(CAG)n$imple_repeat
Seq primer: pyx-5.
FEATURES
Location/Qualifiers
1..732
source
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE: 6832091"
/tissue_type="whole brain"
/dev_stage="embryo 13.5,14.5,16.5,17.5dpc"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH-BMAP_FYO"
/note="Organ: Brain; Vector: pyx-Asc; Site 1: EcoR I;
Site 2: Not I; The library was constructed according
to Bonfido, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured RNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with oligo-dT
primer containing a Not I site. Double strand cDNA was
size selected according to mRNA size fraction, ligated
with EcoR I adaptor, digested with NotI and then cloned
directionally into pyx-Asc vector. The library tag
sequence located between the Not I site and the polyA tail
is AGCGAGACAG. This library was created for the University
of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
Developing Mouse Nervous System', supported by National
Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
program coordinator."
BASE COUNT 197 a 187 c 209 g 139 t
ORIGIN
Query Match 20.0%; Score 727.2; DB 14; Length 732;
Best Local Similarity 99.6%; Pred. No. 7.8e-153;
Matches 729; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 530 CCACATGGCCAAAGAGTTCTTCTTCAAGAGATCTGTCAGGTCACAAATTAAACATTA 589
DB 1 CCACATGGCCAAAGAGTTCTTCTTCAAGAGATCTGTCAGGTCACAAATTAAACATTA 60
QY 590 AGTTACAGAGAGCAATATTCATATGTCGCGAGGCGCCCTCAGACAGAGAAATGGGCTG 649
DB 61 AGTTACAGAGAGCAATATTCATATGTCGCGAGGCGCCCTCAGACAGAGAAATGGGCTG 120
QY 650 TGCCTCAGTCATCAAGGCTGAAATACAGAGATTATTAACAGCAGCAAAACATATGA 709
DB 121 TGCCTCAGTCATCAAGGCTGAAATACAGAGATTATTAACAGCAGCAAAACATATGA 180
QY 710 GTGACACTTAAACAGGTCGCCAGGCAAGAACTATTCTATGCAATGAATTTACCCAGG 769
DB 181 GTGACACTTAAACAGGTCGCCAGGCAAGAACTATTCTATGCAATGAATTTACCCAGG 240
QY 770 CTCAGCTGCTTCAATATGAAATCTTCTGACATGATCAAGATGAAATCTCACTGAG 829
DB 241 CTCAGCTGCTTCAATATGAAATCTTCTGACATGATCAAGATGAAATCTCACTGAG 300
QY 830 AAGAAATTAATCTAGCTATGACCTAATGATGTTGGCATGCTGTGAGCCAGCTGCCG 889
DB 301 AAGAAATTAATCTAGCTATGACCTAATGATGTTGGCATGCTGTGAGCCAGCTGCCG 360
QY 890 CCGTCTGCTTCAAGATATCATCTCTCTTCTTCTTCAAGAGAGTTGCTCCGAGTGGGA 949
DB 361 CCGTCTGCTTCAAGATATCATCTCTCTTCTTCTTCAAGAGAGTTGCTCCGAGTGGGA 420
QY 950 TGTCCGTATAGCTCTTCTTCTGATGATCAAGGCTGCTGAGAGAGCCGTGTAGAGG 1009
DB 421 TGTCCGTATAGCTCTTCTTCTGATGATCAAGGCTGCTGAGAGAGCCGTGTAGAGG 480
QY 1010 ATGACGAGAGCAAGAGAAATGCGCTGTGACATTTGAAGATTAAGAGCGGAGAACT 1069

```

Db	481	ATGAGCAGCAGCCAGAGAAAGAAACCTCGCTGTGACATTTGAAGATATGAAGCGCGGAAC	540
Oy	1070	TCGAGCCGAGCGACAGTGTGTGGAGCTTGGAGAAAGCGCCCGCAAGCCGCTCTTGGAGCAGCAGCGCA	1128
Db	541	TCGAGCCGAGCGGACAGTGTGTGGAGCTTGGAGAAAGCGCCCGCAAGCCGCTCTTGGAGCAGCAGCGCA	600
Oy	1130	AAGAGCAGCAGCAGCGTTGGCTCAGCTGAGACGCGCCGAGCAGAGAGAGAAAGACGCGGAGC	1188
Db	601	AAGAGCAGCAGCAGCGTTGGCTCAGCTGAGACGCGCCGAGCAGAGAGAGAAAGACGCGGAGC	660
Oy	1190	GCCAGAGCAGCAGAGGCGCCAGCGCGCAGCTTGGAGCTTGGAGAAAGCAGTGGAGAAACAGCGGG	1248
Db	661	GCCAGAGCAGCAGAGGCGCCAGCGCGCAGCTTGGAGCTTGGAGAAAGCAGTGGAGAAACAGCGGG	720
Oy	1250	AGCTGGAGCGCGC 1261	
Db	721	AGCTGGAGCGCGC 732	
RESULT 7			
LOCUS	CB248849	725 bp	mRNA linear EST 14-FEB-2003
DEFINITION	UI-M-EXO-py1-m-24-0-UI.r1 NIH_BMAP_EXO Mus musculus cDNA clone		
ACCESSION	CB248849		
VERSION	CB248849.1	GI:28387663	
KEYWORDS	EST.		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclerognathi; Muridae; Murinae; Mus.		
AUTHORS	1 (Baeae 1 to 725)		
TITLE	NIH-MGC http://mgc.nci.nih.gov/ .		
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)		
COMMENT	Unpublished		
	Contact: Robert Strausberg, Ph.D.		
	Email: cgabbs-remail.nih.gov		
	Tissue Procurement: Dr. James Lin, University of Iowa		
	cDNA Library preparation: Dr. M. Bento Soares, University of Iowa		
	cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa		
	DNA Sequencing by: Dr. M. Bento Soares, University of Iowa		
	Clone Distribution: MGC clone distribution information can be		
	found through the I.M.A.G.E. Consortium/LNL at:		
	http://image.llnl.gov		
	This clone was contributed by the Brain Molecular Anatomy Project		
	(BMAP)		
	Seq primer: pyX-5.		

```

FEATURES
SOURCE

Location/Qualifiers
1. 725
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE: 5719103"
/tissue_type="whole brain"
/dev_stage="embryo 15.5 dpc"
/lab_host="DH10B (T1 phage resistant)"
/clone_id="NH_BMAP_EXO"
/note="Organ: brain; Vector: pYX-A5c; Site 1: EcoR I;
Site 2: Not I; The library was constructed according to
Bonaldi, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured mRNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with an
oligo-dT primer containing a Not I site. Double stranded
cDNA was size selected according to mRNA size fraction,
ligated with EcoR I adaptor, digested with Not I, and then
cloned directionally into pYX-A5c vector. The library tag
sequence located between the Not I site and the polyA tail
is GTGCGTGGAA. This library was created for the
University of Iowa Mouse Brain Molecular Anatomy Project
(BMAP): 'Gene Discovery in the Developing Mouse Nervous
System' supported by National Institutes of Mental Health
(NIMH), Hemlin Chin, Ph.D., program coordinator."

```

BASE COUNT	250 a	151 c	221 g	102 t	1 others					
ORIGIN										
Query Match	19.9%; Score 724; DB 14; Length 725;									
Best Local Similarity	99.9%; Pred. No. 4, 1e-152;									
Matches 724; Conservative	0; Mismatches 1; Indels 0; Gaps 0;									
QY	1336	CTTGAATGGGAACCGAACCAGAGACAGAACTCTCTGAATCATAGAGAAACAAGAGACAGAG	1395							
DB	1	CTTGAATGGGAACCGAACCAGAGACAGAACTCTCTGAATCATAGAGAAACAAGAGACAGAG	60							
QY	1396	GGCAACCTGGTCTCGAAGCAAGAGAGAACTCTGAGTTTGAGTTAGTAAGCTCTGAAT	1455							
DB	61	GGCAACCTGGTCTCGAAGCAAGAGAGAACTCTGAGTTTGAGTTAGTAAGCTCTGAAT	120							
QY	1456	GACAAAAAGCATCAGCTTAGAAGAAAACTTCAGATATCAGGTGTCAGCTGGCAACCAG	1515							
DB	121	GACAAAAAGCATCAGCTTAGAAGAAAACTTCAGATATCAGGTGTCAGCTGGCAACCAG	180							
QY	1516	AGGCAAGAATTTGAGACCAAGAACAACTCTAGAGAGTTAAGATTTGTTGAATTCACCCAC	1575							
DB	181	AGGCAAGAATTTGAGACCAAGAACAACTCTAGAGAGTTAAGATTTGTTGAATTCACCCAC	240							
QY	1576	TTACAGCAGCAGCTTGACGAATCTCAGCAATGCTTGAAGAATTATTCAGAGAAACAG	1635							
DB	241	TTACAGCAGCAGCTTGACGAATCTCAGCAATGCTTGAAGAATTATTCAGAGAAACAG	300							
QY	1636	ATACTCAGTGCACAGTTAAAAACAAGTCACAGACAAAGTTTGCATAGAGACTCGCTTCTT	1695							
DB	301	ATACTCAGTGCACAGTTAAAAACAAGTCACAGACAAAGTTTGCATAGAGACTCGCTTCTT	360							
QY	1696	ACCCTCAAAAGAGCTTTGGAAGCAAAAGACTGGCCCCGGACAGACTCCCGGAGACAGCTG	1755							
DB	361	ACCCTCAAAAGAGCTTTGGAAGCAAAAGACTGGCCCCGGACAGACTCCCGGAGACAGCTG	420							
QY	1756	GACGAGGTGGAGAGAGACCAAGCTCAAGCTGAGAGAGATTGATGTTTCAACAACAG	1815							
DB	421	GACGAGGTGGAGAGAGACCAAGCTCAAGCTGAGAGAGATTGATGTTTCAACAACAG	480							
QY	1816	CTGAAGGACTGAGAGAGATACATAGCAAAACAGCACTCCAGAAACAGAGGTCCTGGAG	1875							
DB	481	CTGAAGGACTGAGAGAGATACATAGCAAAACAGCACTCCAGAAACAGAGGTCCTGGAG	540							
QY	1876	GCAGCGGACCTGAAGCAAGAAAGACAGAGAGAGAAAGCTCTGAGTTAGAGAACAAAG	1935							
DB	541	GCAGCGGACCTGAAGCAAGAAAGACAGAGAGAGAAAGCTCTGAGTTAGAGAACAAAG	600							
QY	1936	GAAACGCTCAGAGACGAGTTCAAGAAAGGACAAAGCAATGGCTGGAGCATGTGCACAG	1995							
DB	601	GAAACGCTCAGAGACGAGTTCAAGAAAGGACAAAGCAATGGCTGGAGCATGTGCACAG	660							
QY	1996	GAGAGCAGCAGCAGCCGCCCGGAAACCCACAGAGAGACAGACTGAAGAGGAGAACAGT	2055							
DB	661	GAGAGCAGCAGCAGCCGCCCGGAAACCCACAGAGAGACAGACTGAAGAGGAGAACAGT	720							
QY	2056	GTCAG 2060								
DB	721	GTCAG 725								
RESULT 8										
BM950476										
LOCUS	BM950476	732 bp	mRNA	linear	EST 14-MAR-2002					
DEFINITION	UI-M-BHOP-bun-o-05-0-UI-rl NIH BM95_EHOP Mus musculus cDNA clone									
IMAGE:5687260 5', mRNA sequence.										
ACCESSION	BM950476									
VERSION	BM950476.1	GI:19434066								
KEYWORDS	EST.									
SOURCE	Mus musculus (house mouse)									
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.									
REFERENCE	1 (bases 1 to 732)									

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: Dr. James Lin, University of Iowa
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 This clone was contributed by the Brain Molecular Anatomy Project (BMAP)

FEATURES
 source
 Location/Qualifiers
 1.732
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /clone="IMAGE:5687260"
 /tissue_type="whole brain"
 /dev_stage="embryo 18.5 dpc"
 /lab_host="DH10B (T1 phage resistant)"
 /clone_1ib="NIH BMAP EH0P"
 /note="Organ: brain; Vector: pYX-Asc, Site 1: EcoR I; Site 2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured mRNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with Not I, and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail, is CAGCCACGAC. This library was created for the University of Iowa Mouse Brain Molecular Anatomy Project (BMAP). Gene discovery in the Developing Mouse Nervous System, supported by National Institute of Mental Health (NIMH), Hemin Chin, Ph.D., program coordinator."

BASE COUNT 208 a 172 c 201 g 148 t 3 others
 ORIGIN
 Query Match 19.8%; Score 722.8; DB 12; Length 732;
 Best Local Similarity 99.3%; Pred. No. 7.6e-152;
 Matches 724; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

2817 AGACATGTGTGTTGAGAGAGTTCAAGGTCAAGAGGTTGGTCCCAAGCTTACGT 2876
 4 ACATNNGTGTGTGTTGAGAGAGTTCAAGGTCAAGAGGTTGGTCCCAAGCTTACGT 63
 2877 GAAACTCATTTCAAGGCGCCGTAAGAAATCCACAGCATGATGCGCCCTACTGAAG 2936
 64 GAAACTCATTTCAAGGCGCCGTAAGAAATCCACAGCATGATGCGCCCTACTGAAG 123
 2937 TCCTGTCAAGTCTTAAGAGAGTGTCTCCCGCGCCCAAGCCACTTCCCGAGAGA 2996
 124 TCCTGTCAAGTCTTAAGAGAGTGTCTCCCGCGCCCAAGCCACTTCCCGAGAGA 183
 2997 GTTATATGCGATGATACATACAGAGAGTCTGAGCAAGAGATTAACTTTACAGAGG 3056
 184 GTTATATGCGATGATACATACAGAGAGTCTGAGCAAGAGATTAACTTTACAGAGG 243
 3057 GGATGTGATTTGTGTTACCAAGAAAGATGTGATCGTGAACGGGAAACGGTGGCGCA 3116
 244 GGATGTGATTTGTGTTACCAAGAAAGATGTGATCGTGAACGGGAAACGGTGGCGCA 303
 3117 GTCCGAGGTCTTCCCTCTTAATCTATGTAGAGCTTAAAGATTCAAGAGGCTCTGGAATGC 3176
 304 GTCCGAGGTCTTCCCTCTTAATCTATGTAGAGCTTAAAGATTCAAGAGGCTCTGGAATGC 363
 3177 TGGGAAAAACAGGAGATTAGAAAAAACCTGAATGCGCAGTTATGCTTCTTACGC 3236

Db 364 TGGGAAAAACAGGAGATTAGAAAAAACCTGAATGCGCAGTTATGCTTCTTACGC 423
 Qy 3237 TGCTACTGTGTCCTCCGAACAACCTACCTGCTCTTGCGACGCTGATTCGCGAAAA 3296
 Db 424 TGCTACTGTGTCCTCCGAACAACCTACCTGCTCTTGCGACGCTGATTCGCGAAAA 483
 Qy 3297 GAACCCAGGTGATGTGTGGAAAGAGAACTGCAAGCTCGAGGAAAAACCGCAGTATAG 3356
 Db 484 GAACCCAGGTGATGTGTGGAAAGAGAACTGCAAGCTCGAGGAAAAACCGCAGTATAG 543
 Qy 3357 GTGGTTCCAGCAAAATATGTCAACTTCAAGCCCGGAAACAGCAAAATCAACCCCAAC 3416
 Db 544 GTGGTTCCAGCAAAATATGTCAACTTCAAGCCCGGAAACAGCAAAATCAACCCCAAC 603
 Qy 3417 TGAGCTAACCAAGACCGCAGTGCAGCAGCAGTGTGCGAGTGATCGGATGATGATTA 3476
 Db 604 TGAGCTAACCAAGACCGCAGTGCAGCAGCAGTGTGCGAGTGATCGGATGATGATTA 663
 Qy 3477 CACCGCCAGAACGATGAGAACTAGCCTTCAGCAAAAGCCAGATCATGACGCTCTCA 3536
 Db 664 CACCGCCAGAACGATGAGAACTAGCCTTCAGCAAAAGCCAGATCATGACGCTCTCA 723
 Qy 3537 CAAGAGAGA 3545
 Db 724 CAAGAGAGA 732

RESULT 9
 BM944544 738 bp mRNA linear EST 14-MAR-2002
 LOCUS UI-M-EH0P-pvr-j-08-0-UI_r1 NIH_BMAP_EH0P Mus musculus cDNA clone
 DEFINITION IMAGE:5695975 5', mRNA sequence.
 ACCESSION BM944544 GI:19428129
 VERSION
 KEYWORDS
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus

REFERENCE
 AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 TITLE NIH-MGC <http://mgc.nci.nih.gov/>.
 JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: Dr. James Lin, University of Iowa
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 This clone was contributed by the Brain Molecular Anatomy Project (BMAP)

FEATURES
 source
 Location/Qualifiers
 1.738
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /clone="IMAGE:5695975"
 /tissue_type="whole brain"
 /dev_stage="embryo 18.5 dpc"
 /lab_host="DH10B (T1 phage resistant)"
 /clone_1ib="NIH BMAP EH0P"
 /note="Organ: brain; Vector: pYX-Asc, Site 1: EcoR I; Site 2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured mRNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded

cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with Not I, and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail, is CAGCAGCAGC. This library was created for the 'University of Iowa Mouse Brain Molecular Anatomy Project (BMAP)'. 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemin Chin, Ph.D., program coordinator."

BASE COUNT 198 a 213 c 184 g 143 t

Query Match 19.8%; Score 721.2; DB 12; Length 738;
Best Local Similarity 99.5%; Pred. No. 1.8e-151;
Matches 734; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

```

OY 2310 CCAGACTGAGAGCAGAGATGCTTGAGAGAGCTGAAAGGAGACGGATGTTCCC 2369
DB 1 CCAGACTGAGAGCAGAGATGCTTGAGAGAGCTGAAAGGAGACGGATGTTCCC 60
OY 2370 TGCAGAACTATGAGAGAGAGATTCAGAGAAATAGGTTCCCACTCCAGCCAAACAGTAC 2429
DB 61 TGCAGAACTATGAGAGAGAGATTCAGAGAAATAGGTTCCCACTCCAGCCAAACAGTAC 120
OY 2430 CGATCTGACATCTGCCCCCTGCCCCCAATGAGCTCTGCGTAGAGACCCCTGCTCTTTC 2489
DB 121 CGATCTGACATCTGCCCCCTGCCCCCAATGAGCTCTGCGTAGAGACCCCTGCTCTTTC 180
OY 2490 AGTGAACCTTCTGAGACCCCTCCAGAACCCCAACATGCGGAGACCTTCAGTTCCAGCTG 2549
DB 181 AGTGAACCTTCTGAGACCCCTCCAGAACCCCAACATGCGGAGACCTTCAGTTCCAGCTG 240
OY 2550 GCCCAGCAGCTAAACGAGAGACCGAGAAACGAGCAACTGCGATAGCTGGCGCTCAGCC 2609
DB 241 GCCCAGCAGCTAAACGAGAGACCGAGAAACGAGCAACTGCGATAGCTGGCGCTCAGCC 300
OY 2610 TTCTCTGACCGTACCTAGTGTGGCCAGTTACGGAGAGATAGGCTTTACCCCGACAC 2669
DB 301 TTCTCTGACCGTACCTAGTGTGGCCAGTTACGGAGAGATAGGCTTTACCCCGACAC 360
OY 2670 AGCCACTGAGCTCTCCCAATCTCCGCTCTGGGCGAGGATGAAAGGAGGCTTACA 2729
DB 361 AGCCACTGAGCTCTCCCAATCTCCGCTCTGGGCGAGGATGAAAGGAGGCTTACA 420
OY 2730 AGCCAGAGCCCTGTATCTCTGAGAGCCAAAGAACCAACCTTAATTTTAAAG 2789
DB 421 AGCCAGAGCCCTGTATCTCTGAGAGCCAAAGAACCAACCTTAATTTTAAAG 480
OY 2790 TGACGTCAATCAGCTTGTGAACAGACAGATGTGTGTGTTGGAGAGTTCAAGTCA 2849
DB 481 TGACGTCAATCAGCTTGTGAACAGACAGATGTGTGTGTTGGAGAGTTCAAGTCA 540
OY 2850 GAAAGGTTGGTCCCAAGCTTACGTGAACCTTTCAGGGCCCGTAAGGAAATCCAC 2909
DB 541 GAAAGGTTGGTCCCAAGCTTACGTGAACCTTTCAGGGCCCGTAAGGAAATCCAC 600
OY 2910 AAGCATCGATATGAGCCCTTCTGAGAAAGTCTCTAGTCTTAAAGAGAGTGTCCCGGAC 2969
DB 601 AAGCATCGATATGAGCCCTTCTGAGAAAGTCTCTAGTCTTAAAGAGAGTGTCCCGGAC 660
OY 2970 GCGCAAGCAGCAGCTTCCGAGAGAGAGTTATTCATGATGACATACAGAGAGTTCTGA 3029
DB 661 GCGCAAGCAGCAGCTTCCGAGAGAGAGTTATTCATGATGACATACAGAGAGTTCTGA 719
OY 3030 GCAGAGAGATTTCCTT 3047
DB 720 GCAGAGAGATTTCCTT 737

```

RESULT 10
BU704308 729 bp mRNA linear EST 09-OCT-2002
LOCUS BU704308
DEFINITION UI-M-P00-bzs-j-21-0-UI.r1 NIH_BMAP_P00 Mus musculus cDNA clone

IMAGE: 6406772 5', mRNA sequence.

BU704308

BU704308.1 GI:23632308

EST.

KEYWORDS Mus musculus (house mouse)

SOURCE

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 729)

NIH-MGC Htftp://mgc.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished

CONTACT: Robert Strausberg, Ph.D.

Email: rgs@bpe-remail.nih.gov

Tissue Procurement: Dr. Jim Lin, University of Iowa

cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa

DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LINL at:

http://image.lnl.gov

This clone was contributed by the Brain Molecular Anatomy Project

(BMAP)

Seq primer: pYX-5.

Location/Qualifiers

1..729

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6"

/db_xref="taxon:10090"

/clone="IMAGE: 6406772"

/tissue_type="whole brain"

/dev_stage="embryo 12.5dpc"

/lab_host="DH10B (T1 phage resistant)"

/clone_id="NIH_BMAP_P00"

/notes="Organ: Brain; Vector: pYX-Asc; Site 1: EcoR I;

Site 2: Not I; The library was constructed according

Bonaldi, Lemon and Soares, Genome Research, 6:791-806,

1996. Denatured RNA was size fractionated on a 1% agarose

gel. First strand cDNA synthesis was primed with oligo-dT

primer containing a Not I site. Double strand cDNA was

size selected according to mRNA size fraction, ligated

with EcoR I adaptor, digested with NotI and then cloned

directionally into pYX-Asc vector. The library tag

sequence located between the Not I site and the polyA tail

is TGAGAGAGCC. This library was created for the University

Iowa Brain Anatomy Project (BMAP). 'Gene Discovery in the

Developing Mouse Nervous System', supported by National

Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,

program coordinator."

BASE COUNT 176 a 206 c 174 g 172 t 1 others

ORIGIN

```

Query Match 19.4%; Score 704.8; DB 13; Length 729;  
Best Local Similarity 99.2%; Pred. No. 8.6e-148;  
Matches 719; Conservative 0; Mismatches 3; Indels 3; Gaps 1;  

OY 382 GCTGTGCTCTGTGCGCAATGAGCTCAATTCAGTTGTAATGTCTCCACCTTAGTA 441
DB 1 GCTGTGCTCTGTGCGCAATGAGCTCAATTCAGTTGTAATGTCTCCACCTTAGTA 60
OY 442 TCTTCTGCTCTGAGAGAGAGTCTCCCTGCTTAACGGGGGCTCTCCCGTATCAG 501
DB 61 TCTTCTGCTCTGAGAGAGAGTCTCCCTGCTTAACGGGGGCTCTCCCGTATCAG 120
OY 502 CCTCTGCTGCTTGGCGCATCTGACGACCATGGCCAAAGAGTTCTTCTTCAGAGA 561
DB 121 CCTCTGCTGCTTGGCGCATCTGACGACCATGGCCAAAGAGTTCTTCTTCAGAGA 177
OY 562 TCTGTCCAGGCTACATTAACATTAAGTTACAGAGAGGACATCATTTGATGTGCC 621
DB 178 TCTGTCCAGGCTACATTAACATTAAGTTACAGAGAGGACATCATTTGATGTGCC 237

```


QY 622 AGGCCCCCTCAGCAGAGAGAGTGGCTGTGCTCACTCATCAAGCTGAATATACAGCAG 681
 Db 238 AGGCCCCCTCAGCAGAGAGATGGCTGTGCTCACTCATCAAGCTGAATATACAGCAG 297
 QY 682 TTATTCACAGCAGCAGCAAACTATGATGAGCACTTAAACAGGTCCCAAGCAAGT 741
 Db 298 TTATTCACAGCAGCAGCAAACTATGATGAGCACTTAAACAGGTCCCAAGCAAGT 357
 QY 742 ATTCTCATGATCAATCAAGTTTACCCAGGCTCAGTGTCTCATATGGAATCTTTCTGAC 801
 Db 358 ATTCTCATGATCAATCAAGTTTACCCAGGCTCAGTGTCTCATATGGAATCTTTCTGAC 417
 QY 802 ATTGATCAAGATGAAATCACTGAGAGAAATTTATCTAGTATGACCTAATGAT 861
 Db 418 ATTGATCAAGATGAAATCACTGAGAGAAATTTATCTAGTATGACCTAATGAT 477
 QY 862 GTTGCATGTCTGTGTCAGCAGTCCGCGCTCTGCTCAGAAATACATCCCTCTTC 921
 Db 478 GTTGCATGTCTGTGTCAGCAGTCCGCGCTCTGCTCAGAAATACATCCCTCTTC 537
 QY 922 TTGAGAAAGTTGCTCTCCGCAATGGAGATGCTCCGTCTAATAGCTCTTCTGTGATCAG 981
 Db 538 TTGAGAAAGTTGCTCTCCGCAATGGAGATGCTCCGTCTAATAGCTCTTCTGTGATCAG 597
 QY 982 AGGCTGCTGAGAGAGCCGTGCTGATGAGATGAGCAGCAGCAGAAAGAACTGCTGTG 1041
 Db 598 AGGCTGCTGAGAGAGCCGTGCTGATGAGATGAGCAGCAGCAGAAAGAACTGCTGTG 657
 QY 1042 ACATTTGAAGATTAAGAGCGGAGAACTTTCAGCAGAGCAGTGTGAGCTGAGAAAGCG 1101
 Db 658 ACATTTGAAGATTAAGAGCGGAGAACTTTCAGCAGAGCAGTGTGAGCTGAGAAAGCG 717
 QY 1102 CGCCA 1106
 Db 718 CGCAA 722

RESULT 11
 LOCUS CBS21237 706 bp mRNA linear EST 28-MAR-2003
 DEFINITION UI-M-GHO-cem-g-10-0-UI.r1 NIH BMAP_GHO Mus musculus cDNA clone
 IMAGE: 6841475 5', mRNA sequence.
 ACCESSION CBS21237
 VERSION CBS21237.1 GI:29354592
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NIH-MGC http://mgc.ncl.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished
 Contact: Robert Strausberg, Ph.D.
 Email: rs9abds-remail.nih.gov
 Tissue Procurement: Dr. Jim Lin, University of Iowa
 cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Distribution information can be found at
 http://genome.iowa.edu/distribution/mousefl.html
 This clone was contributed by the Brain Molecular Anatomy Project
 (BMAP).
 The following repetitive elements were found in this cDNA
 sequence: 220-337, >(GGA)n\$simple_repeat
 Seq primer: pyx-5.
 Location/Qualifiers
 1..706
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /clone="IMAGE: 6841475"

FEATURES
 source

/tissue type="Whole brain"
 /dev stage="1, 5, and 15 days newborn"
 /lab_host="DH10B (T1 phage resistant)"
 /clone_id="NIH BMAP_GHO"
 /note="Organ: Brain; Vector: pyx-Asc; Site 1: EcoR I;
 Site 2: Not I; The library was constructed according
 Bonaldi, Lennon and Soares, Genome Research, 6:791-806,
 1996. Denatured RNA was size fractionated on a 1% agarose
 gel. First strand cDNA synthesis was primed with oligo-dT
 primer containing a Not I site. Double strand cDNA was
 size selected according to mRNA size fraction, ligated
 with EcoR I adaptor, digested with NotI and then cloned
 directionally into pyx-Asc vector. The library tag
 sequence located between the Not I site and the polyA tail
 is CGAATCTGAT. This library was created for the University
 Iowa Brain Anatomy Project (BMAP): "Gene Discovery in the
 Developing Mouse Nervous System", supported by National
 Institute of Mental Health (NIMH), Hemlin Chin, Ph.D.,
 program coordinator."

BASE COUNT 219 a 143 c 247 g 96 t 1 others
 ORIGIN
 Query Match 19.3%; Score 701.8; DB 14; Length 706;
 Best Local Similarity 99.6%; Pred. No. 4e-147;
 Matches 703; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 946 GGGATGTCCTGATTAAGCTCTTCTTGTGATCAGAGGCTGCTGAGAGCCGTGTCA 1005
 Db 1 GGGATGTCCTGATTAAGCTCTTCTTGTGATCAGAGGCTGCTGAGAGCCGTGTCA 60
 QY 1006 GAGGATGACAGAGCAGAGCAAGAACTCCCTGTGACATTTGAAGATTAAGAGCGGGAG 1065
 Db 61 GAGGATGACAGAGCAGAGCAAGAACTCCCTGTGACATTTGAAGATTAAGAGCGGGAG 120
 QY 1066 AACTTGACGAGGAGCAGTGTGAGTGAAGAGCGCCGCAAGCGCTCTTGAGCAGCAG 1125
 Db 121 AACTTGACGAGGAGCAGTGTGAGTGAAGAGCGCCGCAAGCGCTCTTGAGCAGCAG 180
 QY 1126 CGCAAGAGCAGAGCGGTTGGCTCAGCTGAGCGCGCCGAGCAGAGAGAGAAAGCGG 1185
 Db 181 CGCAAGAGCAGAGCGGTTGGCTCAGCTGAGCGCGCCGAGCAGAGAGAGAAAGCGG 240
 QY 1186 GAGCGCAGAGAGAGGCGCAAGCGGCGAGCTGAGTGAAGAGCAGCTGAGAGCAG 1245
 Db 241 GAGCGCAGAGAGAGGCGCAAGCGGCGAGCTGAGTGAAGAGCAGCTGAGAGCAG 300
 QY 1246 CGGAGCTGAGAGCGGAGCAGAGAGAGAGAGAGAGATCAGAGCGCGCAGAGCC 1305
 Db 301 CGGAGCTGAGAGCGGAGCAGAGAGAGAGAGAGAGATCAGAGCGCGCAGAGCC 360
 QY 1306 GCAAAACGGAACTGGAAGAGCAGCAACTTTGAATGGGAACGAAACCGAGACAGAA 1365
 Db 361 GCAAAACGGAACTGGAAGAGCAGCAACTTTGAATGGGAACGAAACCGAGACAGAA 420
 QY 1366 CTCCTAATCAAGAGAAACAAGAGCAGAGGGGACCGGTGCTCTGAAGGCAAGAGAGAG 1425
 Db 421 CTCCTAATCAAGAGAAACAAGAGCAGAGGGGACCGGTGCTCTGAAGGCAAGAGAGAG 480
 QY 1426 ACTCTGAGATTGAGTTAGTAAGCTGTAATGACAAAACATCAGCTAGAGAGAACTT 1485
 Db 481 ACTCTGAGATTGAGTTAGTAAGCTGTAATGACAAAACATCAGCTAGAGAGAACTT 540
 QY 1486 CAGGATATCAGGTGTGACTGGCAACCCAGAGCAAGAAATTGAGACAGAAAGTCT 1545
 Db 541 CAGGATATCAGGTGTGACTGGCAACCCAGAGCAAGAAATTGAGACAGAAAGTCT 600
 QY 1546 AGAGACTTAAGATTGCTGAATACCCCTTACAGCAGCAGTGAAGAGATCTGAGCA 1605
 Db 601 AGAGACTTAAGATTGCTGAATACCCCTTACAGCAGCAGTGAAGAGATCTGAGCA 660
 QY 1606 ATGCTTGAAGACTTAATTCAGAGAAACAGATACCTAGTACCACT 1651
 Db 661 ATGCTTGAAGACTTAATTCAGAGAAACAGATACCTAGTACCACT 706


```

/db xref="taxon:10090"
/clone="IMAGE: 6827736"
/tissue_type="whole brain"
/dev stage="embryo 13.5,14.5,16.5,17.5dpc"
/lab host="DH10B (T1 phage resistant)"
/clone_1ib="NIH BMAP FYO"
/notes="Organ: Brain, Vector: pYX-Asc, Site_1: EcoR I;
Site_2: Not I; The library was constructed according
Bonaldi, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured RNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with oligo-dT
primer containing a Not I site. Double strand cDNA was
size selected according to mRNA size fraction, ligated
with EcoR I adaptor, digested with NotI and then cloned
directionally into pYX-Asc vector. The library tag
sequence located between the Not I site and the polyA tail
is AGCGAGACAG. This library was created for the University
Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
Developing Mouse Nervous System', supported by National
Institute of Mental Health (NIMH), Humen Chin, Ph.D.,
program coordinator."

```

Query Match	Similarity	19.0%	Score 691.2	DB 14	Length 713
Best Local	Similarity	98.6%	Pred. No. 9.9e-145		
Matches	704	Conservative	0	Mismatches	9
				Indels	1
				Gaps	1
QY	825	TCGAGAGAATTTATCTAGTATGCACTTAATGATGTTCCATGTCGTGACGCACT	884		
DB	1	TGCAGAAATTTATCTAGTATGCACTTAATGATGTTCCATGTCGTGACGCACT	60		
QY					
DB	885	GCCGCCCTCCTGCTCCAGATACATCCCTCTTCTTCAAGAGATTCCCTCCGCGAG	944		
	61	GCCGCCCTCCTGCTCCAGATACATCCCTCTTCTTCAAGAGATTCCCTCCGCGAG	120		
QY	945	TGGATATGCGCTAATAGCTCTTCTGATGATCAAGAGGCTGCTAGAGCCCTGCTC	1004		
DB	121	TGGATATGCGCTAATAGCTCTTCTGATGATCAAGAGGCTGCTAGAGCCCTGCTC	180		
QY	1005	AGAGATGAGCAGCAGCCAGAGAAACTCCCTGTGATTTTGAAGTAAAGACGGGA	1064		
DB	181	AGAGATGAGCAGCAGCCAGAGAAACTCCCTGTGATTTTGAAGTAAAGACGGGA	240		
QY	1065	GAACTTGCAGCGAGCAGTGTGAGCTGGAAGCGCCGCAAGCGCTCTTGGAGCACA	1124		
DB	241	GAACTTGCAGCGAGCAGTGTGAGCTGGAAGCGCCGCAAGCGCTCTTGGAGCACA	300		
QY	1125	GCGCAAGAGCAGAGACGGTGTGCTCAGCTGAGCGCGCCGACGAGAGAGAAAGCG	1184		
DB	301	GCGCAAGAGCAGAGACGGTGTGCTCAGCTGAGCGCGCCGACGAGAGAGAAAGCG	360		
QY	1185	GGAGCGCCAGAGCAGAGCGCCAGCGCAGCTGAGTGGAGAAAGCACTTGGAGAAACA	1244		
DB	361	GGAGCGCCAGAGCAGAGCGCCAGCGCAGCTGAGTGGAGAAAGCACTTGGAGAAACA	420		
QY	1245	GCGGAGCTGAGCGGCAAGAGAGAGAGAGAGAGAGATTCGAGAGCGCCGAGGC	1304		
DB	421	GCGGAGCTGAGCGGCAAGAGAGAGAGAGAGAGAGATTCGAGAGCGCCGAGGC	480		
QY	1305	CGCAAAACGGGAACCTGAAAGGCAAGCACTTGAATGGGACGGAACCGGAGACAGA	1364		
DB	481	CGCAAAACGGGAACCTGAAAGGCAAGCACTTGAATGGGACGGAACCGGAGACAGA	540		
QY	1365	ACTCTGATATGAGAGAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAA	1424		
DB	541	ACTCTGATATGAGAGAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAA	600		
QY	1425	GACTCTGAGTTGATGATAGAGAGCTCTGAATGACAAAAGCATCAGCTAGAGAGAACT	1484		
DB	601	GACTCTGAGTTGATGATAGAGAGCTCTGAATGACAAAAGCATCAGCTAGAGAG-AAA	659		
QY	1485	TCAGATATCAAGGTCTGACTTGGCAACCCAGAGGCAAGAAATTGAGACAGAA	1538		

Db 660 TCAGGATATCAGGTGTCGACTGGCAACCCANAGGCAAGAATTGAGAGCACGAA 713

RESULT	14
CB244555	
LOCUS	
DEFINITION	688 bp mRNA linear EST 12-FEB-2003
ACCESSION	CB244555
VERSION	UI-M-FY0-cdr-a-13-0-UI.r1 NIH BMAP FYO Mus musculus cDNA clone
KEYWORDS	IMAGE: 6833270 5', mRNA sequence.
SOURCE	CB244555 CB244555.1 GI:28366199
ORGANISM	EST.
	Mus musculus (house mouse)
	Mus musculus

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 688)	NIH-MGC	http://mgc.nci.nih.gov/ .	National Institutes of Health, Mammalian Gene Collection (MGC)	
		Unpublished		
	Contact: Robert Strausberg, Ph.D.			

Tissue Procurement: Dr. Jim Iln, University of Iowa
CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: MGC Clone distribution information can be
found through the I.M.A.G.E. Consortium/BLN at:
<http://image.llnl.gov>

The following repetitive elements were found in this cDNA sequence: 413-530 >(GGA)nhsimple-repeat
Seq primer: pYX-5.

FEATURES	Location/Qualifiers
SOURCE	1. .688

```

/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE: 6833270"
/tissue_type="whole brain"
/dev_stage="embryo.13.5,14.5,16.5,17.5dpc"
/lab_host="DH10B (T1 phage resistant)"
/clone_1lb="NIH BMAP FY0"
/notes="Organ: Brain; Vector: pYX-Asc; Site_1: EcoR I;
Site_2: Not I; The library was constructed according
Bonaldi, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured RNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with oligo-dT
primer containing a Not I site. Double strand cDNA was
size selected according to mRNA size fraction, ligated
with EcoR I adaptor, digested with NotI and then cloned
directionally into pYX-Asc vector. The library tag
sequence located between the Not I site and the poly A tail
is AGCGAGACAG. This library was created for the university
Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
Developing Mouse Nervous System', supported by National
Institute of Mental Health (NIMH), Hemmi Chin, Ph.D.,
program coordinator."

```

	Query Match	18.8%	Score 684.8	DB 14	Length 688
	Best Local Similarity	99.7%	Pred. No. 2.7e-143		
	Matches 666	Conservative	0	Mismatches 2	Indels 0
				Gaps	0
QY	753	ATCAAGTTACCCGAGCTCAGCTGGCTTCATATGCAATCTTCTGCATTTGCATCAGA	8122		
Db	1	ATCAAGTTACCCGAGCTCAGCTGGCTTCATATGCAATCTTCTGCATTTGCATTCACAGA	60		
QY	813	TGGAAGTCACTGCAGAAATTATTCAGTCATGACCACTAATTAATGTTGCATGTC	8727		

```

/mol_type="mRNA"
/db_xref="taxon:9606"
/_clone="IMAGE:618811"
/_sex="male"
/_tissue_type="sympathetic trunk"
/_dev_stage="adult, 16 yr"
/_lab_host="DH10B"
/_clone_id="Lupsk1_sympathetic trunk"
/_note="vector: pCMV-Sport6 (Life Technologies); Site_1:
NotI; Site_2: SalI; cDNA made by oligo-CT priming.
directionally cloned using the following adaptors:
5'-TCGACCCACGCGCTCG-3' and
5'-GACTAGTTCTTGATCGGAGCGCGCCCT(15)-3'. This size selected >
1 kb for average insert length 1.9 kb. This is a primary
library, non-amplified. Library constructed by Life
Technologies and donated by J. Lupsk1, M.D./Ph.D. (Baylor
College of Medicine); available through Life
Technologies."

```

Query Match	18.1%;	Score 658.2;	DB 13;	Length 979;
Best Local Similarity	84.2%;	Pred. No. 3.1e-137;		
Matches	776;	Conservative 0;	Mismatches 141;	Indels 5; Gaps 3;
QY	291	ACTTCCCTCTGTCAATGAAACAGCAACCAATGCTGATTTCCAGTGCACCAAGCATTTGGTAT	350	
DB	1	ACTTCCCTCTGTCAATGAAACAGCAACCAAGTGTGATTTTTCAGCCACAGCATTTGGTAT	60	
QY	351	AGAGAGGATTTGCTAGCATGCCACACATCAAGCGTGTGCTCTCTGTCGAATGGGCTCCAT	410	
DB	61	GGGAGGATATCCCGAGCATATGCCCGCTTACAGCTGTGTCTCAATGTCGAATGGGATTCAT	120	
QY	411	TCCAGTTGTTGGAATGTCTCCACACCTTAATGATCTTCTGTCCTCCACAGCAGAGTGCCTCC	470	
DB	121	TCGAGTTGTTGGAATGTCTCCACACCTTAATGATCTTCTGTTCCACAGAGAGCTGTGCCCC	180	
QY	471	CCTGGCTAACGGGGCTCCTCCCGTATACAGCCTCTGCTGCTGGCTTTGGCATCTCTCAGC	530	
DB	181	CCTGGCTAACGGGGCTCCTCCCGTATATCAACCTCTGCTGCTGCTGCTGCTGCTGCTGCTG	240	
QY	531	CACATGGCCAAAGAGTCTTCTCTCAGCAATCTGGTCCAGGGTCCACAAATTAAACCTAA	590	
DB	241	CACATTTGCCAAAGAGTCTTCTCTTATGATCTGGTCCAGGGTCCACAACTTAAACCTAA	300	
QY	591	GTTACAGAGGACCAATCATTCAGTGTGCGCAGCGCCCTCCACAGCAGAGATGGGCTGT	650	
DB	301	ATTACAAAGGACCAATCATTTGATGTGCGCAGTGTCCACAGTGGCAGAGTGGGCTGT	360	
QY	651	GCTCAGTCATCAAGGCTGGAATACAGGCAGTTATTCACAGCCACGACAAACTATATAG	710	
DB	361	TCTCAGTCATCAAGGCTGGAATACAGGCAGTTATTCATATGATGATGACAAACTATATAG	420	
QY	711	TGACACTTAAACAGTCCCCCAGGCAAGACTATTTCTCATGCAATCAAGTTTACCACAGC	770	
DB	421	TGACACTTAAACAGTCCCCCAGGCAAGACTATTTCTATGCAATCAAGTTTACCACAGG	480	
QY	771	TCAGTGGCTCAATATGGAATCTTTTGAATTTGATGATCAAGATGGAATCATCATGCGAGA	830	
DB	481	TCAGTGGCTCAATATGGAATCTTTTGAATTTGATGATGGAATGGAATCTTACGCGAGA	540	
QY	831	AGAAATTATTCAGTATGACACTTAATTTGATGTTGGCATGTCTGGTCAGCCATGCGCC	890	
DB	541	GAATTTATTCAGTATGACACTTAATTTGATGTTGGCATGTCTGGTCAGCCATGCGCC	600	
QY	891	CGTCTGCTTCAAGATATCATTCCTCTTCTTCTTCAAGAGTTGCTCTCGGATGGGAT	950	
DB	601	TGCTCTGCTTCAAGATATCATTCCTCTTCTTCTTCAAGAGTTGATCTGCGAGTGGAT	660	
QY	951	GTCGCTAATAGCTCTTCTTCTGAGATCAAGGCTGCTGAGGAGCCGTGTCAGAGGA	1010	
DB	661	ATCTGCTAATAGCTCTTCTTCTGAGATCAAGGCTGCTGAGGAGGAAGGATTTTGAAGA	720	

QY 1011 TGAGCAGCAGC---CAGAGAGAGAACTGCTGTGACATTGAG-ATAGAGCGGAG 1066
 Db 721 TGAACAACAATATTAGAAAAGAAATTACCTGTACGTTGAGATTAAGAGCGGAG 780
 QY 1067 ACTTCGAGCGAGGCACTGTGAGCTGAGAAAGCGCCCAAGCGCTCTTGAGCAGCAGC 1126
 Db 781 ACTTTGACGTCGCAACCTGGAAGTGAAGAAAGAG-CAAGCTCTCTGAGAAAGCGCAGC 839
 QY 1127 GCAAGAGCAGAGCGGTTGCTCAGCTGAGAGCGCGCGAGCAGAGAGAGAAAGCGCG 1186
 Db 840 GCAAGAGAGCAGAGCGCTGCGCCANNCTGAACGCGCGAGCCAGAGAAAGAGAAAGT 899
 QY 1187 AGCGCCAGAGCAGAGAGCGCAA 1208
 Db 900 GACCCCAAGAACAGAGCGCAA 921

Search completed: December 5, 2003, 00:39:55
 Job time : 4658.38 secs

THIS PAGE BLANK (USPTO)


```

Db      402 ATGCCAATCTGTCAATTCATTCAGCAATTCGCTCAGTTGACCTATGCAACACCCCTG 461
Qy      401 -----TGGGCTCAATTCAGTTGTTGGAATGTCTCAACCTTATATCT 444
Db      462 TCTTCTGTACTTTCAGGACCAATATCTCCCTCAATGATGCTGCTCCCTATGCTCT 521
Qy      445 TCTGTCCCTCAGCAGCAGTGCCTCCCTGGCTAAACGGGGCTCCCTCCGCTATACAGCT 504
Db      522 TCTGTAGTACATCTCTCAATTCACCAATGAAATGCGCATCTCATTACGCTTTATCACT 581
Qy      505 CTGCTGTGCTTGGCATCTGACGCCACATGCGCCAAAGATTCCTTCAGCAGATCT 564
Db      582 CTTATTTCTTCTTCAACATTCGCTCAATGATCATCTTACAGCTGATGATGGAGATTT 641
Qy      565 GGTCCAGGTCACAACTTAAACATTAAGTTACGAAGGCAATCACTTGTATGCGCAGC 624
Db      642 GGTGTGTAGTATTCAGAAAGGCCCAATCTGTATGATTTAGATTCATAGTCAACT 701
Qy      625 GCGCCCTCCAG-----CAGCAGAAATGGGCTGTG 651
Db      702 TCTTCAATCTGCTTCCCTCTCAGGAACTCACCCTAAGACAGGACCTCAGATGGCAGTT 761
Qy      652 CCTCAGTCAAGGCTGAAATACAGGAGTTATTCACAGCCACGACAAACTATAGAT 711
Db      762 CCTCAGCTTCAAGTTAAATATGCGCAAAATTTAATAGCTTACGAAGAGCATGAGC 821
Qy      712 GACACTTAACAGGTCCCGAGCAAACTATTCTCATGCAATCAAGTTTACCCAGCT 771
Db      822 GGAATCCTCTCAGGTTTCAAGCTAAGAAATGCCCTTCTCAGTCAAACTCTCTCAACT 881
Qy      772 CAGTGGCTTCAATATGATCTTTCTGACATTCAGATTCAGATGAAATCACTGACAGAA 831
Db      882 CAGCTAGCTTACTATTGGAATCTGTGCTCAATTCATGATGATGACGACATTTGAAAGCTGA 941
Qy      832 GAATTTATCTAGCTATGACCTAATTTGATGTTGTCATGTCAGCCCACTGCGGCC 891
Db      942 GAATTTATCTGCGGAGTCACCTCATCTGACATGCGCAAGCTGACGCGCACTACATG 1001
Qy      892 GTCTGCTCTCCAGAAATATCCCTCTCTTCCTTACAGAAATGTCCTCCGCGCATG 951
Db      1002 ACGTTGCTCTCCGAGCTGTGCTCTCCTCATTCTTACAGAGGGGAAAGCAAG----- 1050
Qy      952 TCCGCTAATAGCTCTTCTTCTGTGATCAGAGGCTGCTGAGGCGCTGCTGACAGGAT 1011
Db      1051 -----TTGATTTCTGTTATGAACTCTGCTTCAATTCAGAAACACAGAA 1097
Qy      1012 GAGCAGCAGCAGAGAGAACTGCTGTGACATTTGAAATTAAGAAACGGGAGATTC 1071
Db      1098 GA---AGAGCTTCAGAAAGAACTGCAATTAATTTGAGAACAAACGAAAGCCAACTAT 1154
Qy      1072 GAGGAGGCACTGTGAGCTGGAAGAGCGCGCCAGCGCTCTTGGAGCAGCAGCGCAA 1131
Db      1155 GAAAGAGAGAAATATGAGACTGGAAGAGGACGCGCAAGTGTGATGAGACACAGAGAGG 1214
Qy      1132 GAGCAGAGCGGTTGGCTCAGCTGAGCGCGCCGAGCAGAGAGGAAAGAGCGGAGGCGC 1191
Db      1215 GAGGCTGAACCGAAAGCCCAAGAAAGAAAGAGTGGAGCGGAAACAGAGAACTTG 1274
Qy      1192 CAGAGCAGAGAGCGCAAGCGCAGCTGAGAGCTGAGAGCAGCTGAGAGACAGCGGAG 1251
Db      1275 CAAGAGCAAGATGAGAAAGACAGCTGAGATTGAGAAACCTTGGAGAAACAGAGAGAG 1334
Qy      1252 CTGAGACCGGACGAGAGAGAGAGAGAGAAAGATCGAGAGCGCGCGCAAAA 1311
Db      1335 CTGAGAGACAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1394
Qy      1312 CGGGAATCTGAAAGGACAGCACTTGAATGGAGAGAGAGAGAGAGAGAGAGAGAGAG 1371
Db      1395 CAGAGCTTGGAGAGCAAGCGCTTTAGAAATGGAGAGAGAGAGAGAGAGAGAGAGAG 1454
Qy      1372 AATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1431

```

```

Db      1455 AGTCAGAGACAGAGAGAAACAGAAAGACATTGTACAGCTGAGCTCCAGAAAGAAATGCTC 1514
Qy      1432 GAGTTTATGATTAAGAGCTCTGAATGACAAAGAAAGATACGCTAGAGAGAAATTCAGAGAT 1491
Db      1515 CACTGGAACCTGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1574
Qy      1492 ATCAGTGTGCACTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1551
Db      1575 GTCCAAATCAGAAAGCAACAAACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1634
Qy      1552 CTAAAGATTCGTGAATCAATCCCACTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1611
Db      1635 CTGGAATTTATGAGAAATCAAACTTCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1694
Qy      1612 GGAAGACTTATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1671
Db      1695 ATCTATCTGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1754
Qy      1672 AGTTGATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1731
Db      1755 AACACACTGATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1814
Qy      1732 CGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1791
Db      1815 TGCCAAAGACTTAAAGAAATTAAGATGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAG 1874
Qy      1792 GAGATTTATGTTTCAACCACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1851
Db      1875 GAAATGATTCATTTAAACATCAGCTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1934
Qy      1852 CTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1911
Db      1935 TTAGCCCTTGAACAATCTCAATTAATCAAAAGTGAACAAATTAAGAGAGAGAGAGAG 1994
Qy      1912 AGCTGAGATTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1934
Db      1995 AGATTAGAGCAAAAAAAAAAAAAA 2017

```

```

RESULT 2
US-09-338-933-72
; Sequence 72, Application US/09338933
; Patent No. 6488931
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer Lynn
; APPLICANT: King, Gordon B.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY OF
; FILE REFERENCE: 210121.462C1
; CURRENT APPLICATION NUMBER: US/09/338,933
; NUMBER OF SEQ ID NOS: 312
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 72
; LENGTH: 2017
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-338-933-72

```

```

Query Match      13.1%; Score 477.4; DB 4; Length 2017;
Best Local Similarity 55.5%; Pred. No. 1.9e-109;
Matches 1111; Conservative 0; Mismatches 796; Indels 96; Gaps 6;

Qy      1 ATGGCTAGTTTCCCAACCTTTTGGTGTAGAGCTGTGATGTCTGGGCTATATCTGTGAG 60
Db      42 ATGGCTAGTTTCCCAACGAGATGAATGAGGGCAAAATATGTGGCTATATCAATCTGAA 101
Qy      61 GAAAGGCGCAAGATGACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
Db      102 GAAAGTATTAAGATGATTAACAGTTTGAATCACTCAACCTTCAGAGAGTTACATACA 161
Qy      121 GGTATCAAGCAGAGAACTTTTTCATCTGGGTTACTGACCTGTCTTAGCAAA 180

```

Db 162 GGTATCAAGCCGTACTTTTCTTACAGTCAAGTCTCCGCGCCCGGTTTACGTGA 221
 Oy 181 ATATGGCGCTAGCGGACATGAAATTAAGTGAAGGATGATCAAGTGAATTTTCATA 240
 Db 222 AATATGGCGCTTATCAGATCTGAACAGATGGAGATGAGCCAGCAAGATGCTTATA 281
 Oy 241 GCCATGAAGCTTATCAAACTGAAGCTACAAAGATATCAGTCCCTCCCACTTCCCTC 300
 Db 282 GCTATGAATCTATCAAGTTAAAGTTGACGGGCAACAGCTGCTGTAGTCCCTCCT 341
 Oy 301 GTATGAACAGCAACAGT---GCTATTTTCCAGTGCACAGCATTTGTATGAGAG 357
 Db 342 ATCAAGAAACAAACCCCTATGTTCTCTCACTATCTGCTGTTTGGATGGAGAC 401
 Oy 358 ATTGCTAGCATG-----CCACACTCAAGCTGTTGCTCTGTGCA----- 400
 Db 402 ATGCCCAATCTGCTCACTTATCAAGCATGCTCCAGTTGACCTTATGACCAACCTT 461
 Oy 401 -----TGGGCTCCATTCAGTTGTGATGTCTCCACCTTATGATCT 444
 Db 462 TCTTCTGCTACTTACAGGACCAAGTATTCCTCCCTAATATGCTGCTCCCTAGTGC 521
 Oy 445 TCTGTCCCTCCAGCAGCAGTGCCTCCCTGCTTACGGGCTCTCCCTCATACAGCT 504
 Db 522 TCTGTATGATCACTCTCATTTACCAATGGAACGCTCATTCAGCTTATCAT 581
 Oy 505 CTGCTGCTTGGCGCATCTCTGACGACATGGCCAAAGATTTCTTCCCTTACAGCAT 564
 Db 582 CTTATTTCTTCTTCAACATTTGCTCATGATCATCTTACAGCTGATGAGAGATTT 641
 Oy 565 GGTCCAGGCTCACAATTAACAATAATTACAGAGGCACAATCATTCATGCTGCGCAG 624
 Db 642 GGTGGGTAGATCCAGAAAGGCCAGTCTGTGATTTAGATCTAGTACCTA 701
 Oy 625 GCCCTCCAG-----CAGCAGATGGGCTGTG 651
 Db 702 TCTTAACCTGCTTCCCTCTCAGGAACTCACTAAGACAGGACCTCAAGTGGGAGT 761
 Oy 652 CCTCATCTCAAGGCTGAATACAGGCACTTATCAAGCAGCACAAACTATGAT 711
 Db 762 CCTCAGCTTCAAGATTAAGTATCGGCAAAATTTAATGCTTACAAAGGCACTGAC 821
 Oy 712 GGAATTAACAGATCTCCAGGCAAGATCTATCTCATGCAATCAAGTTTACCCAGCT 771
 Db 822 GGAATCTCTCAGTCTTCAAGTGAATATGCCCTTCTTCAATCAATCTCTCTCAAT 881
 Oy 772 CAGCTGCTTCAATATGAAATCTTCTGACATTTGATCAAGATGAAACTCATCTGCA 831
 Db 882 CAGCTAGTACTATTTGACTCTGGCTGACATGATGATGAGGACAGTTGAAGCTGA 941
 Oy 832 GAATTTATCTAGCTATGACCTAATTTGATTTGCTGATGCTGATGCTGCTGCGCC 891
 Db 942 GAATTTATCTGCGATGACCTCATGACATGGCCAAAGCTGACAGCACCTACACATG 1001
 Oy 892 GTCTGCTCCAGAAATATCCCTCTTCTTCAAGAGATTCGCTCCGCGAGTGGATG 951
 Db 1002 AGTGTGCTCCGAGGCTGTCTCTCATTTTCAAGGGGGAAGAG----- 1050
 Oy 952 TCCGCTATAGCTCTTCTTCTGATCAGAGGCTGCTGAGAGCCCTGCTCAGAGAT 1011
 Db 1051 -----TTGATTTCTGTATGAACTCTGCTTCAATCAGAAAAACAGAA 1097
 Oy 1012 GAGCAGACCCAGAGAAAGAACTGCTGTGACATTTGAAGATTAAGAGCGGAGAACTTC 1071
 Db 1098 GA---AGAGCTCAGAGAAAGCTGCAATTTTGAAGACAAACGAAACCACTAT 1154
 Oy 1072 GAGCAGAGCAGTGTGAGCTGAGAGAGCGCCAGAGCCCTCTGAGCAGACGCAAA 1131
 Db 1155 GAAAGAGAAACATGAGCTGAGAGAGCAGCAGCTGATGATGAGCAGCAGAGAG 1214
 Oy 1132 GAGCAGAGCAGTGTGAGCTGAGAGCGCCGAGCAGAGAGAGAGAGCGGAGCGC 1191
 Db 1215 GAGGCTGAAGCAAGCCAGAAAGAGAGAGAGTGGAGGAGGAGAAACAGAGAACTG 1274

Oy 1192 CAGAGCAGAGAGCCAGCGGAGCTGAGCTGAGAGAGCAGCTGAGAGAGCGGAG 1251
 Db 1275 CAGAGCAGAGAGTGAAGAGAGCTGAGTGTGAGAAACGTTGAGAAACAGAGAG 1334
 Oy 1252 CTGAGCGGACAGAGAGAGAGAGAGAGATGAGAGCGGAGCGGCAAA 1311
 Db 1335 CTGAGAGAGCAGCGGAGAGAGAGAGAGAGATGAGAAAGACAGAGCAGCAAAA 1394
 Oy 1312 CGGAACTGGAAGGACAGCACTTGAATGGAAACGAAACCGAGACAGAACTCTG 1371
 Db 1395 CAGAGCTTGAAGACAAACGCTTTAGAAATGGGAAACCTCTGCGAGAGCTGCTC 1454
 Oy 1372 AATCAGAGAAACAGAGAGAGAGCGGACCGGCTCTGAGAGCAAGAGAGAGCTG 1431
 Db 1455 ACTCAGAGAACAGAGAAACAGAGACATTTGACAGCTGAGCTCCAGAAAGAAAGCTC 1514
 Oy 1432 GAGTTGATTAAGAGCTGATGATGACAAAGATCAGCTAGAGAGAAACCTCAGAT 1491
 Db 1515 CACCTGGAAGCTGAGAGAGATGAGAAACATCAGCAGATCTCAGGAGACTACAGAT 1574
 Oy 1492 ATCAGGTGCACTGGCAACCAAGAGCAAGAAATTGAGAGCAGACAGCTTGAAG 1551
 Db 1575 GTCCAAATCAGAAACCAACCAAGAGCTGAGAGAGTTTGGATPAACAGTGTGAC 1634
 Oy 1552 CTAAAGATTGCTGAATACACCACTTACAGCAGAGCTGAGAGATCTCAGCAATGCT 1611
 Db 1635 CTGAAATTAAGAAATCAACACTTCAACAGAGCTTAAAGAAATCAAAATTAAGCT 1694
 Oy 1612 GGAAGACTTAATCCAGAGAAACAGATCTCAGTACAGTGAACAGTTAAACAAGTCCAGAGAAC 1671
 Db 1695 ATCTATCTGTCTCCGAGAGAGAGATTAATTAAGAAATTAATTAATTAAGTCACT 1754
 Oy 1672 AGTTGATGAGAGCTCTCTTCTTACCTCAAGAGCTTGAAGCAAGAGCTGGC 1731
 Db 1755 AACACCTGATTCAGGATCAGTTACTTAAAGATCAAGAAAGAGAAATTA 1814
 Oy 1732 CGGAGCAGCTCCGAGAGAGCTGAGAGAGTGAAGAGAGAGAGAGAGAGAGTGAAG 1791
 Db 1815 TGCCAAAGCTTAAGAAACATTAATGATCTCTTGAAGAAAGAGAGCTGAAGCTTCA 1874
 Oy 1792 GAGATTAATGTTTCAACAAACAGCTGAGAGATCAGAGAGATCATAGCAACAGCA 1851
 Db 1875 GAAATGATTAATTAAGATGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAG 1934
 Oy 1852 CTCAGAGAGAGAGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1911
 Db 1935 TTAGCTTGAACACTTCATTAATCAAGCTGACAAATGAGAGAAATGAGAAAGAA 1994
 Oy 1912 AGCTGAGATTAAGAGAGAGAG 1934
 Db 1995 AGATTAGAGCAAAAAA 2017

RESULT 3
 US-09-215-681-72
 Sequence 72, Application US/09215681A
 Patent No. 6528253
 GENERAL INFORMATION:
 APPLICANT: Mitcham, Jennifer L.
 APPLICANT: Frudakis, Tony N.
 APPLICANT: King, Gordon E.
 TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSIS
 TITLE OF INVENTION: OF OVARIAN CANCER
 FILE REFERENCE: 210121.463
 CURRENT APPLICATION NUMBER: US/09/215,681A
 CURRENT FILING DATE: 1998-12-17
 NUMBER OF SEQ ID NOS: 310
 SOFTWARE: faSeq for Windows Version 3.0
 SEQ ID NO 72
 LENGTH: 2017
 TYPE: DNA
 ORGANISM: Homo sapien

US-09-215-681-72

Query Match 13.1%; Score 477.4; DB 4; Length 2017;
Best Local Similarity 55.5%; Pred. No. 1.9e-109;
Matches 1111; Conservative 0; Mismatches 796; Indels 96; Gaps 6;

1 ATGGCTAGTTTCCACACCTTTGGGTAGCTGATGTCTGGGCCATTAATCTGTGAG 60
42 ATGGCTAGTTTCCACACCTTTGGGTAGCTGATGTCTGGGCCATTAATCTGTGAG 101
61 GAAAGGCGCAAGCATGACAGAGTTCCTTAGCCTGAAGCCGATAGCGGATTTATTA 120
102 GAACGCTACTAAGCATGATTAACAGTTTGATTAACCTCAACCTTAGAGGATTAACA 161
121 GGTATCAAGGAGAACCTTTTTCATCTGGGTACTGACCTGCTCTTGACCA 180
162 GGTATCAAGGCGGCTACTTTTCTTACAGTGAAGTCTGGCGCCCGTTTATGCTGA 221
181 ATATGGGCGTAGCGGACATGATTAACGATGAAGATGATCAAGTGAATTTTCCATA 240
222 ATATGGGCGTTATCATGATCTGAACAGATGGAAGATGACACCAAGATTTCTTA 281
241 GCGATGAAGCTTATCAAACTGAAGCTACAGATATCAGCTTCCCTCCACATTTCCCT 300
282 GCTATGAAGCTATCAAGTAAAGTTGAGGCGCAACAGCTGCTGATGCTCCCTCT 341
301 GTCATGAAGCAACCAAGT---GCTATTTCCAGTGAACAGATTTGGTATGAGAG 357
342 ATCATGAAGCAACCCCTATGTTCTTCCACTAATCTTGTCTGTTTGGAGATGGAAG 401
358 ATATGGTACATG---CCACACCTACAGCTGTTGCTCTGCTGCTCA--- 400
402 ATGGCCATCTGTTCATTCATGACCATTTGCTCCAGTTGACCTATAGCAACACCTTG 461
401 -----TGGGCTCATTCACGTTGTGGAATGTCTCAACCTTATGATCT 444
462 TCTTGTGCTACTTACGAGGACAGATATCTCCCTTAATGATGCTGCTCCCTAGTGTCT 521
445 TCTGTCCCTTCCAGCAGAGTGTCTCCCTGGCTAACGGGGCTCTCCCTCATACAGCT 504
522 TCTGTATGATACATCTCATTAACCAATGAAATGACAGTCTCATTCACCTTTATCCAT 581
505 CTGCTGTGCTTGGCATCTGTCAGCAGCATGAGCAAGAGTCTTCTTCCAGCATCT 564
582 CTTATTTCTTCTTCAACATGCTCTATGATCTTACAGCTGATGATGAGAGATTT 641
565 GGTCCAGGATCACAATTAACACTAAGTTACAGAAAGCAACATCTTGAATGTGCCAGC 624
642 GGTGTGTAGTATCCAGAAAGGCCAGTCTGATTGATTAGATCTAGTAGTCACT 701
625 GGCCTCCAG-----CAGCAGATGGGCTGTG 651
702 TCTCTAACCTGCTTCCCTCTCAGGGAACCTCACTTAAGACAGGACCTCAGAGTGGCAGT 761
652 CTTGATCATCAAGGCTGAATATACAGGAGTATTAACAGCAGCAACAAATCTATAGT 711
762 CTTGAGCTTCAAGATTAAGTATGCGCAAAATTTAATAGTCTGACAAAGGAGCTAGC 821
712 GAGACATTAACAGGTCCCGAGCAAGAACTATTTCTATGACATCAAGTTTACCCAGGCT 771
822 GGTATCTCTCAGGTTTCAAGCTAAGAAATGCTTCTTCACTCAATCTCTCAAACT 881
772 CAGCTGGCTTCAATATGGAATTTTCTGACATTTGATCAAGATGAAATCTACTGCAAA 831
882 CAGCTAGCTATCAATTTGAGCTCTGCTGATCAATGATGAGAGCAGTGAAGCTTAA 941
832 GAAATTTATCTAGCTATGACCTAATTTGATGATGATGATGATGATGATGATGATGAT 891
942 GAAATTTATCTAGCTATGACCTAATTTGATGATGATGATGATGATGATGATGATGAT 1001
892 GTCTGCTCTCCAGAAATACATCCCTCTCTCTTCCAGAAAGATTCGCTCCGAGAGTGG 951
1002 AGCTTGGCTCCCGAGCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1050

952 TCCGTATTAAGCTCTTCTCTGTGATCAGAGGCTGCTGAGAGCCGCTGTCAGAGAT 1011
1051 -----TTGATTTCTGTTATGAACTCTGCTCTTCAATATCAGAAACCAAGAA 1097
1012 GAGCAGCAGCAGAGAAAGAAATGCTGCTGATCAATTTGAATTAAGAACGGGAACTTC 1071
1098 GA---AAGCCTCAGAAAGAAATGCTGCTGATCAATTTGATGAGCAAAAGGAAAGCTAT 1154
1072 GAGCAGCAGCAGTGTGAGCTGAGAAAGCCGCTCAAGCCTCTTGTGAGCAGCAGCGAAA 1131
1155 GAACGAGAAACATGAGCTGAGAAAGCCGCAAGTGTGATGAGCAGCAGCAGAGG 1214
1132 GAGCAGCAGCAGTGTGCTGATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1191
1215 GAGCCTAAGCAGAAAGCCGCAAGAAAGAGAAAGATGAGAGCGGAAACAGAGAACTG 1274
1192 CAGAGCAGAGAGCCAGAGCAGCAGCTGAGAGAGCAGCTGAGAGAGCAGCGGAG 1251
1275 CAGAGCAGAGATGAGAAAGAGCAGCTGAGAGAGAAAGCTTGAAGAAACAGAGAG 1334
1252 CTGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGATCGAGAGCGCGAGCGCAAAA 1311
1335 CTGAGAGAGAGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1394
1312 CCGGAACTGGAAGAGCAGCAGCACTTGAATGGGAAACGAAACCGGAGACAGAACTCTG 1371
1395 CAGAGCTTGAAGAGCAACCGCTTTGAATGGGAAAGATCTGCTGCGAGAGCTGCTC 1454
1372 AATCAGAGAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1431
1455 AGTCAGAAAGCAAGGAGAAAGAAAGACATTTGAGCTGAGCTCAGAAAGAAAGCTTC 1514
1432 GAGTTGAGTTAAGAGCTCTGAATGACAAAGATAGCTAGAGAGAAATTTTCAAGAT 1491
1515 CACCTGAACTGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1574
1492 ATCAGTGTGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1551
1575 GTCCAAATCAGAAAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1634
1552 CTAAAGATGCTGAATACCCACTTACAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1611
1635 CTGGAATTTATGAAATCAACCACTTCAACAGAGCTTAAAGAAATTAACATTAAGCT 1694
1612 GGAAGATTTATTCAGAGAAACAGATCTAGTGAACAGAGTTAAAGAGCTCCAGCAGAAC 1671
1695 ATCTATCTGTGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1754
1672 AGTTGATAGAGAGCTGCTTCTTAACTCAAAAGAGCTTGGAGAGAGAGAGAGAGAG 1731
1755 AACACCTGATTCAGAGATCAGTTTACTTATTAAGATCATCAGAAAGAGAGAGATTA 1814
1732 CCGCAGAGAGCTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1791
1815 TGCCAAAGAGCTTAAAGAAACATTAAGATGATCTTGAAGAAAGAGAGAGAGAGAG 1874
1792 GAGATGATGTTTCAACCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1851
1875 GAAATGATTCATTAACATTCAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1934
1852 CTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1911
1935 TTAGCCTTGAACAATCTTAAATTAACAGTGAACAAATTTGAAGAAATTCAGAAAGAAA 1994
1912 AGCTGAGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1934
1995 AGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2017

RESULT 4
US-08-630-915A-193
; Sequence 193, Application US/08630915A

Patent No. 6309820
GENERAL INFORMATION:
APPLICANT: SPARKS, Andrew B.
APPLICANT: HOFFMAN, No. 6309820h
APPLICANT: KAY, Brian K.
APPLICANT: FOWLES, Dana M.
APPLICANT: McCONNELL, Stephen J.
TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
NUMBER OF SEQUENCES: 227
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/630,915A
FILING DATE: 03-APR-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Mirock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-174
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 193:
SEQUENCE CHARACTERISTICS:
LENGTH: 2873 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: DNA
US-08-630-915A-193

Query Match 9.1%; Score 332.4; DB 4; Length 2873;
Best Local Similarity 61.0%; Pred. No. 3.8e-73;
Matches 631; Conservative 0; Mismatches 361; Indels 42; Gaps 4;

QY 2609 CTTCTCTGACCGTACCTAGTGTGCGCAAGTTACGCGAGATCAGCCTTACCCCGCA 2668
DB 541 CTTTTCACAACTAATCTAATATCATGCGAGAAATACGCTTCACTCGAATCG 600
QY 2669 CAGCCACTGCTCTCCCATCTCCCGCTCGGCGCCAGGGGTGAAAGGTGAAGGGCTAC 2728
DB 601 TGTCCCTG--ATCTGATACCTATTCATGACAGGAGCAAGGTGAGAAACTTAA 657
QY 2729 AAGCGCAAGCCCTGTATCCCTGAGAGCCAAAGAACACCACTTAATTTTACAAA 2788
DB 658 AAGCAAGAGCCCTTGTCTTCTGAGTGAAGAAATATGAGTGAAGTGAAGTGAAG 717
QY 2789 GTGAGCTACACCGTCTGGAACAGACAGATGTGTGTTGGAGAGTTCAAGTTC 2848
DB 718 ATGACATTATCTGCTTGGAGACGAAAGAAATTTGGGGAGGTGATGAG 777
QY 2849 AGAAGGTTGTTCCCAAGTCTTACGTAAATCTATTCAAGGCGCCGTAAAGAAATCA 2908
DB 778 GAGAGGATGTTTCCCAATCTTATGTCAAGTCAATCTCGGAGATGAAGTAAACGGG 837
QY 2909 CAAGCATGATCTGCGCTTACGAAAGTCTGTGTTAAAGAGAGTGGCTTCCCGGG 2968
DB 838 AA-----GAACAGAAAGCTTTGTATGACCTGTAAATGAAGAACTTACTCGG 885
QY 2969 CCGCAAGCAGCATTCCTCGGAGAAAGATTATTCATGTACATACGAGAGTTCTG 3028

DB 886 CAGCCATTCAGTT-----GGAGAGAAATATATGCACTTATTCATATTCAGAGTGG 939
QY 3029 AGCAAGAGATTTAACTTTACGCAAGGGAGATGTATGTGTACCAAGAAAGATGTCG 3088
DB 940 AACCTGAGATTTGCTTTCACAGAAAGTGAAGAAATATGTGTGACCCAGAAAGATGAG 999
QY 3089 ACTGTGAGCGGGAACGGTGGCGCAAGTCCGAGTCTTCCCTTAATCTATGTAGAGGC 3148
DB 1000 AGTGTGACAGAAATGATTTGAGATGAGATGAGATGAGATTTTTCATCAAACTATGTAAC 1059
QY 3149 TTAAGATTACAGAGGCTCTGGAACCTGAGGAAACAGAGGATTTAGAGAAAAAAGCTG 3208
DB 1060 CAAGGATTAAGAGATTTTGGAGATGCTAGCAAGTGTGACATCAATTAATTAACCTG 1119
QY 3209 AATTGCCAGGTTATGCTTCTTACAGCTGTCTACTGTGTCCGAACAACCTACCTGCTC 3268
DB 1120 AGATTGCTCAGGTAATTCAGCATATGTTGCTTGTGTGAAACAATTAACCTTGAC 1179
QY 3269 CTGGGAGCTGATTCGATCCGGAAGAAAGAACCCAGGTGAGTGGAGAGAACTGCG 3328
DB 1180 CAGGACAGTTAATTAATTTTAAAGAAATATACAGAGGTGTGCGCAAGAGATTC 1239
QY 3329 AAGCTGAGGGAAGAAAGCGCCAGATAGGTGTTCAGCAATTAATGTCAACTTCTAA 3388
DB 1240 AGGCGAGAGGAAAGAAAGCAAGAAAGATGTTTCTGCCAGTCAATGTTAACTTTGG 1299
QY 3389 GCCCGGAACAAAGCAAAATACCCCACTAGCTACCCAAAGCCGAGTGCAGCCAGCAG 3448
DB 1300 GTCAAGTAGTGAAGAGCCACACCTGCTTTCATC-----CTG 1338
QY 3449 TGTGCAAGTGTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3508
DB 1339 TATGTGAGTGTATCTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1398
QY 3509 GCAAGGCGCAGATCATCAAGCTCTCAACAGAGAGACCCGAGTGTGAGAAAGAGAG 3568
DB 1399 CCAAGGACAACTATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1458
QY 3569 TGAATGGGCAAGTGGCTCTTCCATTCATTAATGATGATGATGATGATGATGATGATG 3628
DB 1459 TCAACGGGATGACTGATCTTCTTCTTCAACTAGTTAAGATGACGACAGACTCAGATC 1518
QY 3629 CAGCCAGCAATGA 3642
DB 1519 CAAGTCAACAGTGA 1532

RESULT 5
US-08-630-915A-39
Sequence 39, Application US/08630915A
Patent No. 6309820
GENERAL INFORMATION:
APPLICANT: SPARKS, Andrew B.
APPLICANT: HOFFMAN, No. 6309820h
APPLICANT: KAY, Brian K.
APPLICANT: FOWLES, Dana M.
APPLICANT: McCONNELL, Stephen J.
TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
NUMBER OF SEQUENCES: 227
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA: US/08/630,915A
 APPLICATION NUMBER: US/08/630,915A
 FILING DATE: 03-APR-1996
 CLASSIFICATION: 536
 ATTORNEY/AGENT INFORMATION:
 NAME: Mirock, S. Leslie
 REGISTRATION NUMBER: 18,872
 REFERENCE/DOCKET NUMBER: 1101-174
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 790-9090
 TELEFAX: (212) 869-8864/9741
 INFORMATION FOR SEQ ID NO: 39:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 747 bases
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA
 US-08-630-915A-39

Query Match 7.1%; Score 258; DB 4; Length 747;
 Best Local Similarity 63.9%; Pred. No. 7.6e-55;
 Matches 418; Conservative 0; Mismatches 215; Indels 21; Gaps 1;
 Db 2989 GGAGAGGATTATTGTCATGATACATACAGAGTTCTGAGCAGAGGATTTAACTTT 3048
 115 GGAGAGGATTATTGTCATGATACATACAGAGTTCTGAGCAGAGGATTTAACTTT 174
 3049 CAGAGAGGAGTGTGATTGTGGTTACGAAGAGATGCTGCTGAGACGGGAAACGGTG 3108
 175 ACAGAGGAGTGTGATTGTGGTTACGAAGAGATGCTGCTGAGACGGGAAACGGTG 234
 3109 GCGCAGAGTGTGATTGTGGTTACGAAGAGATGCTGCTGAGACGGGAAACGGTG 3168
 235 GGAGAGTGTGATTGTGGTTACGAAGAGATGCTGCTGAGACGGGAAACGGTG 294
 3169 GGAATGCTGGGAAACAGGAGTTTAAAGAAAAAAGCTGAAATTTGCCAGTTTGTCT 3228
 295 GGAATGCTGGGAAACAGGAGTTTAAAGAAAAAAGCTGAAATTTGCCAGTTTGTCT 354
 3229 TCCCTAGTGTGATTGTGGTTACGAAGAGATGCTGCTGAGACGGGAAACGGTG 3288
 335 GCAATGTTGTTCTGTTCTGTTCTGTTCTGTTCTGTTCTGTTCTGTTCTGTTCT 414
 3289 CCGAAGAGTGTGATTGTGGTTACGAAGAGATGCTGCTGAGACGGGAAACGGTG 3348
 415 CTAAGAGAGTGTGATTGTGGTTACGAAGAGATGCTGCTGAGACGGGAAACGGTG 474
 3349 CAGATAGGTTGTTCTGTTCTGTTCTGTTCTGTTCTGTTCTGTTCTGTTCTGTT 3408
 475 CAGAGAGTGTGATTGTGGTTACGAAGAGATGCTGCTGAGACGGGAAACGGTG 534
 3409 ACCCACTGAGTGTGATTGTGGTTACGAAGAGATGCTGCTGAGACGGGAAACGGTG 3468
 535 AACCTGCTTTCACTCT-----GTAATGAGTGTGTTCTGTTCTGTTCTGTTCT 573
 3469 TAGAGTTACACCGCCAGAGATGAGTGTGTTCTGAGCAAGAGGCGAGATCATCAAC 3528
 574 TATGCTAGTGTGTTCTGTTCTGTTCTGTTCTGTTCTGTTCTGTTCTGTTCTGTT 633
 3529 GTCTCTCAACAGAGAGACCGGAGTGTGTTCTGTTCTGTTCTGTTCTGTTCTGTT 3588
 634 GTTATGAACAAAGATGATCTGTTCTGTTCTGTTCTGTTCTGTTCTGTTCTGTT 693
 3589 TTCCATCAATTAATGTAAGTGTGAGTGTGTTCTGTTCTGTTCTGTTCTGTTCTGTT 3642
 694 TTCTCTCAATTAATGTAAGTGTGAGTGTGTTCTGTTCTGTTCTGTTCTGTTCTGTT 747

RESULT 6
 US-09-404-879A-60

Sequence 60, Application US/09404879A
 Patent No. 648846
 GENERAL INFORMATION:
 APPLICANT: Mitcham, Jennifer L.
 APPLICANT: King, Gordon E.
 APPLICANT: Algate, Paul A.
 TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
 TITLE OF INVENTION: DIAGNOSIS OF OVARIAN CANCER
 FILE REFERENCE: 210121.462C1
 CURRENT APPLICATION NUMBER: US/09/404,879A
 CURRENT FILING DATE: 1999-09-24
 NUMBER OF SEQ ID NOS: 393
 SOFTWARE: FastSeq for Windows Version 3.0
 SEQ ID NO: 60
 LENGTH: 480
 TYPE: DNA
 ORGANISM: Homo sapien
 US-09-404-879A-60

Query Match 4.0%; Score 144; DB 4; Length 480;
 Best Local Similarity 62.1%; Pred. No. 1.9e-26;
 Matches 247; Conservative 0; Mismatches 145; Indels 6; Gaps 1;
 Db 1 ATGGCTCAGTTTCCACACCTTTCGTTGTTAGCTGATGCTGAGCATTAATCTGTGAG 60
 14 ATGGCTCAGTTTCCACACCTTTCGTTGTTAGCTGATGCTGAGCATTAATCTGTGAG 73
 61 GAAGAGGCAAGATGACACAGCTTCTTACCTGAGACCGGATTTATTAAT 120
 74 GACGTTCTAAGTGTGTTAAACGTTGATTAACCTTCAAGAGGTTACATACA 133
 121 GGTATCAAGAGAGTGTGTTTTCCTCAATCTGTTTACCTGAGCTGTTTACGACAA 180
 134 GGTATCAAGAGAGTGTGTTTTCCTCAATCTGTTTACCTGAGCTGTTTACGACAA 193
 181 ATATGGGCGTACGAGCATGAATACATGAGATGAGATGATGAGATTTTCATA 240
 194 ATATGGGCGTACGAGCATGAATACATGAGATGAGATGAGATTTTCATA 253
 241 GCAATGAGTGTGTTTTCCTCAATCTGTTTACCTGAGACCGGATTTATTAAT 300
 254 GGTATCAAGAGAGTGTGTTTTCCTCAATCTGTTTACCTGAGCTGTTTACGACAA 313
 301 GTCATGAACAGCAACAGAGTGTGTTTTCCTCAATCTGTTTACCTGAGCTGTTTAC 360
 314 ATCAAGAAACACCC-----CTATGTTCTTCACTAATCTGCTGTTTGGATG 367
 361 GCTAGATGCCACCATCAGCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 398
 368 GGAAGCATGCCCATCTGTCATTCATCAGCATTGCC 405

RESULT 7
 US-09-338-933-60
 Sequence 60, Application US/09338933
 Patent No. 6488931
 GENERAL INFORMATION:
 APPLICANT: Mitcham, Jennifer Lynn
 APPLICANT: King, Gordon E.
 TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY OF
 TITLE OF INVENTION: OVARIAN CANCER
 FILE REFERENCE: 210121.462C1
 CURRENT APPLICATION NUMBER: US/09/338,933
 CURRENT FILING DATE: 1999-06-23
 NUMBER OF SEQ ID NOS: 312
 SOFTWARE: FastSeq for Windows Version 3.0
 SEQ ID NO: 60
 LENGTH: 480
 TYPE: DNA
 ORGANISM: Homo sapien
 US-09-338-933-60

Query Match 4.0%; Score 144; DB 4; Length 480;

Best Local Similarity 62.1%; Pred. No. 1.9e-26;
Matches 247; Conservative 0; Mismatches 145; Indels 6; Gaps 1;

```
OY 1 ATGGCTCAAGTTTCCCAACCTTTCGCTGAGTGTGATGCTGGGCCAATACCTGGAG 60
    |||
DB 14 ATGGCTCAAGTTTCCCAACCTTTCGCTGAGTGTGATGCTGGGCCAATACCTGGAG 73
    |||
OY 61 GAAAGGGCCAGAGATGACGAGCAGTTCCTTAGCCTGAGCCGATAGCGGATTTATTA 120
    |||
DB 74 GAAAGGCTAGAGATGATTAACAGTTTGAATTAACCTTAACCTTACGAGATTAATTA 133
    |||
OY 121 GGTGATCAAGCAGAGAACTTTTTCATCTGGGTACCTCAGCCTGTCTTAGACAA 180
    |||
DB 134 GGTGATCAAGCAGCCTTCTTTTCTACAGTGTCTGCGGCCCGGCTTTAGCTGAA 193
    |||
OY 181 ATATGGCCCTAGCCGACATGAAATACATGGAAGATGATCAATGGAATTTTCA 240
    |||
DB 194 ATATGGCCCTTATTCATATCTGAAACAGATGGAGATGAGACCAAGAGTTCTTATA 253
    |||
OY 241 GCCATGAAGCTTATCAACTGAAGTACAGATATACGCTCCCTCCACACTTCCCT 300
    |||
DB 254 GCTATGAAGCTTATCAAGTAAAGTTGACAGGCCCAACAGCTGCTAGTCTTCTCT 313
    |||
OY 301 GTCATGAACAGCAACAGCTGCTATTTCCAGTGCACCAAGATTTGATAGAGGAT 360
    |||
DB 314 ATCATGAACCAACCC-----CTATGTTCTCTCCACTATCTCTGCTTTGGGATG 367
    |||
OY 361 GGTAGATGCCACCACTACAGCTGTGCTCTGTCGCC 398
    |||
DB 368 GGAGCATGCCCAATCTGTTCATTCATCAGCAATGCC 405
    |||
```

RESULT 8

US-09-215-681-60
; Sequence 60, Application US/09215681A
; Patent No. 6528253
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Frudakis, Tony N.
; APPLICANT: King, Gordon E.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSIS
; FILE REFERENCE: 210121.463
; CURRENT APPLICATION NUMBER: US/09/215,681A
; CURRENT FILING DATE: 1998-12-17
; NUMBER OF SEQ ID NOS: 310
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 60
; LENGTH: 480
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-215-681-60

Query Match 4.0%; Score 144; DB 4; Length 480;
Best Local Similarity 62.1%; Pred. No. 1.9e-26;

Matches 247; Conservative 0; Mismatches 145; Indels 6; Gaps 1;

```
OY 1 ATGGCTCAAGTTTCCCAACCTTTCGCTGAGTGTGATGCTGGGCCAATACCTGGAG 60
    |||
DB 14 ATGGCTCAAGTTTCCCAACCTTTCGCTGAGTGTGATGCTGGGCCAATACCTGGAG 73
    |||
OY 61 GAAAGGGCCAGAGATGACGAGCAGTTCCTTAGCCTGAGCCGATAGCGGATTTATTA 120
    |||
DB 74 GAAAGGCTAGAGATGATTAACAGTTTGAATTAACCTTAACCTTACGAGATTAATTA 133
    |||
OY 121 GGTGATCAAGCAGAGAACTTTTTCATCTGGGTACCTCAGCCTGTCTTAGACAA 180
    |||
DB 134 GGTGATCAAGCAGCCTTCTTTTCTACAGTGTCTGCGGCCCGGCTTTAGCTGAA 193
    |||
OY 181 ATATGGCCCTAGCCGACATGAAATACATGGAAGATGATCAATGGAATTTTCA 240
    |||
DB 194 ATATGGCCCTTATTCATATCTGAAACAGATGGAGATGAGACCAAGAGTTCTTATA 253
    |||
```

```
OY 241 GCCATGAAGCTTATCAACTGAAGTATACAGATATACGCTCCCTCCACACTTCCCT 300
    |||
DB 254 GCTATGAAGCTTATCAAGTAAAGTTGACAGGCCCAACAGCTGCTAGTCTTCTCT 313
    |||
OY 301 GTCATGAACAGCAACAGCTGCTATTTCCAGTGCACCAAGATTTGATAGAGGAT 360
    |||
DB 314 ATCATGAACCAACCC-----CTATGTTCTCTCCACTATCTCTGCTTTGGGATG 367
    |||
OY 361 GGTAGATGCCACCACTACAGCTGTGCTCTGTCGCC 398
    |||
DB 368 GGAGCATGCCCAATCTGTTCATTCATCAGCAATGCC 405
    |||
```

RESULT 9

US-09-404-879A-5
; Sequence 5, Application US/09404879A
; Patent No. 6468546
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: King, Gordon E.
; APPLICANT: Algate, Paul A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.462C2
; CURRENT APPLICATION NUMBER: US/09/404,879A
; CURRENT FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 393
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 531
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-404-879A-5

Query Match 4.0%; Score 144; DB 4; Length 531;
Best Local Similarity 62.1%; Pred. No. 2e-26;
Matches 247; Conservative 0; Mismatches 145; Indels 6; Gaps 1;

```
OY 1 ATGGCTCAAGTTTCCCAACCTTTCGCTGAGTGTGATGCTGGGCCAATACCTGGAG 60
    |||
DB 42 ATGGCTCAAGTTTCCCAACCTTTCGCTGAGTGTGATGCTGGGCCAATACCTGGAG 101
    |||
OY 61 GAAAGGGCCAGAGATGACGAGCAGTTCCTTAGCCTGAGCCGATAGCGGATTTATTA 120
    |||
DB 102 GAAAGGCTAGAGATGATTAACAGTTTGAATTAACCTTAACCTTACGAGATTAATTA 161
    |||
OY 121 GGTGATCAAGCAGAGAACTTTTTCATCTGGGTACCTCAGCCTGTCTTAGACAA 180
    |||
DB 162 GGTGATCAAGCAGCCTTCTTTTCTACAGTGTCTGCGGCCCGGCTTTAGCTGAA 221
    |||
OY 181 ATATGGCCCTAGCCGACATGAAATACATGGAAGATGATCAAGTGAATTTTCA 240
    |||
DB 222 ATATGGCCCTTATTCATATCTGAAACAGATGGAGATGAGACCAAGAGTTCTTATA 281
    |||
OY 241 GCCATGAAGCTTATCAACTGAAGTATACAGATATACGCTCCCTCCACACTTCCCT 300
    |||
DB 262 GCTATGAAGCTTATCAAGTAAAGTTGACAGGCCCAACAGCTGCTGATGCTCTCT 341
    |||
OY 301 GTCATGAACAGCAACAGCTGCTATTTCCAGTGCACCAAGATTTGATAGAGGAT 360
    |||
DB 342 ATCATGAACCAACCC-----CTATGTTCTCTCCACTATCTCTGCTTTGGGATG 395
    |||
OY 361 GGTAGATGCCACCACTACAGCTGTGCTCTGTCGCC 398
    |||
DB 396 GGAGCATGCCCAATCTGTTCATTCATCAGCAATGCC 433
    |||
```

RESULT 10

US-09-338-933-5
; Sequence 5, Application US/09338933
; Patent No. 6488931
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer Lynn

APPLICANT: King, Gordon E.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY OF
TITLE OF INVENTION: OVARIAN CANCER
FILE REFERENCE: 210121.462C1
CURRENT APPLICATION NUMBER: US/09/338.933
CURRENT FILING DATE: 1999-06-23
NUMBER OF SEQ ID NOS: 312
SOFTWARE: FASTSEQ for Windows Version 3.0
SEQ ID NO 5
LENGTH: 531
TYPE: DNA
ORGANISM: Homo sapien
US-09-338-933-5

Query Match 4.0%; Score 144; DB 4; Length 531;
Best Local Similarity 62.1%; Pred. No. 2e-26;
Matches 247; Conservative 0; Mismatches 145; Indels 6; Gaps 1;

QY 1 ATGGCTCAGTTTCCACACCTTTGCGTGTAGCTGTGAGCCATTAAGTGTGAG 60
DB 42 ATGGCTCAGTTTCCACACCTTTGCGTGTAGCTGTGAGCCATTAAGTGTGAG 101
QY 61 GAAAGGCCAAGCATGACGACGAGTTCCTTGAAGCCGATGAGCGGATTTACT 120
DB 102 GAAAGGCCAAGCATGACGAGTTCCTTGAAGCCGATGAGCGGATTTACT 161
QY 121 GGTATCAAGGAGAACTTTTTCCTTGAAGCCGATGAGCGGATTTACT 180
DB 162 GGTATCAAGGAGAACTTTTTCCTTGAAGCCGATGAGCGGATTTACT 221
QY 181 ATATGGCGCTTACGACATGATTAAGATGATGATGATGATGATGATGATGAT 240
DB 222 ATATGGCGCTTACGACATGATTAAGATGATGATGATGATGATGATGATGAT 281
QY 241 GCCATGAAGCTTATCAACTGAAGCTGAAGATGATGATGATGATGATGATGAT 300
DB 282 GCTATGAAGCTTATCAACTGAAGCTGAAGATGATGATGATGATGATGATGAT 341
QY 301 GTCATGAAGCAAGCAAGCTGCTTTCCTGATGACGATGATGATGATGATGATGAT 360
DB 342 ATCATGAAGCAAGCAAGCTGCTTTCCTGATGACGATGATGATGATGATGATGAT 395
QY 361 GCTAGCATGCCACCATCTACAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 398
DB 396 GGAAGCATGCCCAATCTGCTCATTCATTCATTCATTCATTCATTCATTCATTC 433

RESULT 11
US-09-215-681-5
Sequence 5, Application US/09215681A
Patent No. 6528253
GENERAL INFORMATION:
APPLICANT: Mitcham, Jennifer L.
APPLICANT: Frudakis, Tony N.
APPLICANT: King, Gordon E.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSIS
TITLE OF INVENTION: OF OVARIAN CANCER
FILE REFERENCE: 210121.463
CURRENT APPLICATION NUMBER: US/09/215.681A
CURRENT FILING DATE: 1998-12-17
NUMBER OF SEQ ID NOS: 310
SOFTWARE: FASTSEQ for Windows Version 3.0
SEQ ID NO 5
LENGTH: 531
TYPE: DNA
ORGANISM: Homo sapien
US-09-215-681-5

Query Match 4.0%; Score 144; DB 4; Length 531;
Best Local Similarity 62.1%; Pred. No. 2e-26;
Matches 247; Conservative 0; Mismatches 145; Indels 6; Gaps 1;
QY 1 ATGGCTCAGTTTCCACACCTTTGCGTGTAGCTGTGAGCCATTAAGTGTGAG 60

DB 42 ATGGCTCAGTTTCCACACGATGAATGAGGCGCAATATGTGGCTATTACATCTGA 101
QY 61 GAAAGGCCAAGCATGACGACGAGTTCCTTGAAGCCGATGAGCGGATTTACT 120
DB 102 GAAAGGCCAAGCATGACGAGTTCCTTGAAGCCGATGAGCGGATTTACT 161
QY 121 GGTATCAAGGAGAACTTTTTCCTTGAAGCCGATGAGCGGATTTACT 180
DB 162 GGTATCAAGGAGAACTTTTTCCTTGAAGCCGATGAGCGGATTTACT 221
QY 181 ATATGGCGCTTACGACATGATTAAGATGATGATGATGATGATGATGATGAT 240
DB 222 ATATGGCGCTTACGACATGATTAAGATGATGATGATGATGATGATGATGAT 281
QY 241 GCCATGAAGCTTATCAACTGAAGCTGAAGATGATGATGATGATGATGATGAT 300
DB 282 GCTATGAAGCTTATCAACTGAAGCTGAAGATGATGATGATGATGATGATGAT 341
QY 301 GTCATGAAGCAAGCAAGCTGCTTTCCTGATGACGATGATGATGATGATGATGAT 360
DB 342 ATCATGAAGCAAGCAAGCTGCTTTCCTGATGACGATGATGATGATGATGATGAT 395
QY 361 GCTAGCATGCCACCATCTACAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 398
DB 396 GGAAGCATGCCCAATCTGCTCATTCATTCATTCATTCATTCATTCATTCATTC 433

RESULT 12
US-08-728-323A-1
Sequence 1, Application US/08728323A
Patent No. 5948676
GENERAL INFORMATION:
APPLICANT: Chang, Yuan
APPLICANT: Bohenzky, Roy A.
APPLICANT: Russo, James J.
APPLICANT: Edelman, Isidore S.
APPLICANT: Moore, Patrick S.
TITLE OF INVENTION: Immediate Early Protein From Kaposi's
TITLE OF INVENTION: Sarcoma-Associated Herpesvirus, DNA
TITLE OF INVENTION: Encoding Same And Uses Thereof
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS: 21
ADDRESS: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/728.323A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 0575/52268/JPM/MS/SKS
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-278-0400
TELEFAX: 212-391-0525
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3489 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:

NAME/KEY: CDS
LOCATION: 1..3489
US-08-728-323A-1

Query Match 4.0%; Score 144; DB 2; Length 3489;
Best Local Similarity 45.1%; Pred. No. 5e-26;
Matches 534; Conservative 0; Mismatches 650; Indels 0; Gaps 0;

```

QY 975 GATCAGAGGCTGCTGAGAGCCGCTGCTGAGAGTGAAGACACAGAGAAAGAACT 1034
DB 1578 GCGAGGACCAACAGCAGAGAGCCACAGCAGAGAGCCACAGCAGAGAGCCACAGCA 1637
QY 1035 GCCTGTGACATTTGAAGATTAAGAGCGGAGAACTTGAAGAGAGAGAGTGTGAGCTGA 1094
DB 1638 GCGAGGACCAACAGCAGAGAGCCACAGCAGAGAGCCACAGCAGAGAGCCACAGCA 1697
QY 1095 GAAAGCCGCGCCAGAGCCCTCTTGAGAGCAGCAGCAGCAGAGAGAGAGAGCTGCTGAGCT 1154
DB 1698 GCGAGGACCAACAGCAGAGAGCCACAGCAGAGAGCCACAGCAGAGAGCCACAGCA 1757
QY 1155 GAGGCGCGCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1214
DB 1758 GCGAGGACCAACAGCAGAGAGCCACAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1817
QY 1215 GCTGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1274
DB 1818 GCGAGGATGAGCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1877
QY 1275 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1334
DB 1878 GCGAGGATGAGCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1937
QY 1335 ACTTGAAATGGGAAACGGAACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1394
DB 1938 GCGAGGATGAGCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1997
QY 1395 GCGGACCGCTGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1454
DB 1998 GCGAGGATGAGCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2057
QY 1455 TGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1514
DB 2058 GCGAGGATGAGCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2117
QY 1515 GAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1574
DB 2118 GCGAGGATGAGCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2177
QY 1575 CTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1634
DB 2178 GCGAGGATGAGCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2237
QY 1635 GATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1694
DB 2238 GCGAGGATGAGCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2297
QY 1695 TACCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1754
DB 2298 GCGAGGATGAGCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2357
QY 1755 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1814
DB 2358 GCGAGGATGAGCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2417
QY 1815 GCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1874
DB 2418 GCGAGGATGAGCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2477
QY 1875 GCGAGGATGAGCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1934
DB 2478 GCGAGGATGAGCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2537
QY 1935 GGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1994

```

```

DB 2538 GCAAGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2597
QY 1995 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2054
DB 2598 GCGAGGATGAGCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2657
QY 2055 TGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2114
DB 2658 GCGAGGATGAGCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2717
QY 2115 GCTTTCCATCCGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2158
DB 2718 GGAAGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2761

```

RESULT 13

```

US-09-298-568-1
; Sequence 1, Application US/09298568
; Patent No. 6322792
; GENERAL INFORMATION:
; APPLICANT: Kieff, Elliott D.
; APPLICANT: Ballestae, Mary E.
; APPLICANT: Kaye, Kenneth M.
; TITLE OF INVENTION: RHADINO VIRUS LANA ACTS IN TRANS ON A UNIT OF RHADINO
; FILE REFERENCE: 16412-10001R
; CURRENT APPLICATION NUMBER: US/09/298,568
; EARLIER FILING DATE: 1999-04-21
; EARLIER APPLICATION NUMBER: US 60/109,422
; EARLIER FILING DATE: 1998-11-19
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 3489
; TYPE: DNA
; ORGANISM: Kapoli's sarcoma-associated herpesvirus
US-09-298-568-1

```

Query Match 4.0%; Score 144; DB 4; Length 3489;
Best Local Similarity 45.1%; Pred. No. 5e-26;
Matches 534; Conservative 0; Mismatches 650; Indels 0; Gaps 0;

```

QY 975 GATCAGAGGCTGCTGAGAGCCGCTGCTGAGAGTGAAGACAGAGAGAGAGAGAACT 1034
DB 1578 GCGAGGACCAACAGCAGAGAGCCACAGCAGAGAGCCACAGCAGAGAGCCACAGCA 1637
QY 1035 GCCTGTGACATTTGAAGATTAAGAGCGGAGAACTTGAAGAGAGAGAGAGAGAGAGAGAGAG 1094
DB 1638 GCGAGGATGAGCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1697
QY 1095 GAAAGCCGCGCCAGAGCCCTCTTGAGAGCAGCAGCAGCAGAGAGAGAGAGAGAGAGAGAGAG 1154
DB 1698 GCGGAGGACCAACAGCAGAGAGCCACAGCAGAGAGCCACAGCAGAGAGCCACAGCA 1757
QY 1155 GAGGCGCGCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1214
DB 1758 GCGGAGGACCAACAGCAGAGAGCCACAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1817
QY 1215 GCTGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1274
DB 1818 GCGAGGATGAGCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1877
QY 1275 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1334
DB 1878 GCGAGGATGAGCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1937
QY 1335 ACTTGAAATGGGAAACGGAACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1394
DB 1938 GCGAGGATGAGCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1997
QY 1935 GCGGACCGCTGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1454

```

Db 1998 GCAGGATGACGACGACGAGATGAGCAGAGGATGAGCAGCAGATGAGCAGCA 2057
Qy 1455 TGCACAAAAGCATCTAGCTAGAGAAAATTTCAGAGATTCAGAGTTCAGCTGGCAGCA 1514
Db 2058 GCAGATGACGACGACGAGATGAGCAGCAGAGATGAGCAGCAGAGATGAGCA 2117
Qy 1515 GAGCAGAAAATTGAGACACGAACTGCTAGAGCTTGAATTCACCA 1574
Db 2118 GAGACGACGATGAGCAGACGACGAGATGAGCAGCAGATGAGCAGCAGCA 2177
Qy 1575 CTTCAGCAGCAGCTTGCAGAACTTCAGCAATGCTTGAAGATTATTCAGAGAAA 1634
Db 2178 GAGATGACGACGACGAGATGAGCAGCAGAGATGAGCAGCAGAGATGAG 2237
Qy 1635 GATCTCAGTACCTGATTAATTAATCAAGTCCAGCAGAACTTTCATAGACTGCTTCT 1694
Db 2238 GCAGCAGCAGCAGATGAGCAGCAGCAGCAGAGAGAGCAGCAGCAGAGAGAG 2297
Qy 1695 TACCTCAAAAAGAGCTTGAAGCAAGAGCTGCGCAGCAGCTCCGAGCAGCACT 1754
Db 2298 GCAGAGATGAGAGCAGCAGCAGAGATGAGAGATGAGAGCAGAGATGAGAGCA 2357
Qy 1755 GACAGAGGTGAGAGAGACCAAGTCAAGCTGAGAGATGATGTTTTCACACCA 1814
Db 2358 GAGCAGAGATGAGAGCAGCAGCAGAGATGAGAGCAGCAGAGATGAGAGCA 2417
Qy 1815 GCTGAAGAACTGAGAGATATCATGCAACAGCACTCCAGAGCAGAGTCCCTTGA 1874
Db 2418 GCAGAGCAGAGATGAGAGCAGCAGCAGAGATGAGAGCAGCAGAGATGAG 2477
Qy 1875 GGCAGCGGACCTGAGCAGAAAGCAGAGAGAGAGCTGAGATGAGAGCAAAA 1934
Db 2478 GAGCAGCAGAGCAGAGATGAGAGCAGCAGAGATGAGAGCAGAGATGAGAG 2537
Qy 1935 GGAAGACCTCAGACAGAGATTCAGAGAAAGGACAGCAATGCTGAGCATGTCAGCA 1994
Db 2538 GCAAGAGCAGAGATGAGAGCAGCAGAGAGAGAGAGAGAGATGAGAGAGT 2597
Qy 1995 GAGAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2054
Db 2598 GAGAGAGCAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2657
Qy 2055 TGTCAAGAAAGAGCGGAGAGAGAGAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2114
Db 2658 GCAGAGAGCAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATGAGAGAGT 2717
Qy 2115 GCTTTTCATCCGATCAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2158
Db 2718 GGAAGAGCAGAGCAGCAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2761

RESULT 14
US-09-410-399-1
; Sequence 1, Application US/09410399
; Patent No. 6482587
; GENERAL INFORMATION:
; APPLICANT: Robertson, Eile S.
; TITLE OF INVENTION: Methods to Inhibit or Enhance the Binding of Viral DNA
; FILE REFERENCE: UM-03778
; CURRENT APPLICATION NUMBER: US/09/410,399
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 3489
; TYPE: DNA
; ORGANISM: Kapoisi's sarcoma-associated herpesvirus
US-09-410-399-1

Query Match 4.0%; Score 144; DB 4; Length 3489;
Best Local Similarity 45.1%; Pred. No. 5e-26;

Matches 534; Conservative 0; Mismatches 650; Indels 0; Gaps 0;
Qy 975 GATTCAGAGGCTGCTTGAAGAGCCGCTGTCAGAGATGAGCAGAGCAGAGAAACT 1034
Db 1578 GCAGAGCCCAACAGCAGAGCAGCAGCAGAGCAGAGCAGCAGAGCAGAGCAGCA 1637
Qy 1035 GCTCTGACATTTGAAATAGAAGCCGAGAACTTTCAGCAGAGCAGTGTGAGCTGGA 1094
Db 1638 GCAGAGCCCAACAGCAGAGGAGAGCCACAGCAGCAGAGCCCAAGCAGAGGAGCCACAGCA 1697
Qy 1095 GAAGCCGCCAGCAGCTCTTGAAGCAGCAGCCCAAGAGCAGAGAGCTGCTCAGCT 1154
Db 1698 GCGGAGCCCAACAGCAGAGGAGCCACAGCAGCAGCCCAAGCAGAGGAGCCACAGCA 1757
Qy 1155 GAGCCGCGCAGAGAGAGAGAAAGCCGAGAGCCGACAGAGCAGAGAGCCAGCGCA 1214
Db 1758 GCGGAGCCCAACAGCAGAGGAGCCACAGCAGCAGAGATGAGCAGCAGAGATGAGCAGCA 1817
Qy 1215 GCTGAGCTGAGAGAGCAGCTGAGAGAGCAGCAGCAGCTGAGAGCAGCAGAGAGAG 1274
Db 1818 GCAGAGATGAGAGCAGCAGAGATGAGCAGCAGAGATGAGCAGCAGAGATGAGCAGCA 1877
Qy 1275 GAGGAGAGAGAGATCAGAGAGCCGAGCCGCAAAACGAGAACTGAGAAAGCAGCA 1334
Db 1878 GCAGAGATGAGCAGCAGAGATGAGCAGCAGAGATGAGCAGCAGAGATGAGCAGCA 1937
Qy 1335 ACTTGAATGGAAGAGAAACCGAGACAGAGAACTCTGATTCAGAGAAACAGAGAGAG 1394
Db 1938 GCAGAGATGAGCAGCAGAGATGAGCAGCAGAGATGAGCAGCAGAGATGAGCAGCA 1997
Qy 1395 GCGCAGCTGCTTGAAGCAGCAGAGAGAGAGAGCTCTGAGATGAGTGAAGAGCTGAA 1454
Db 1998 GCAGAGATGAGAGCAGCAGAGATGAGCAGCAGAGATGAGCAGCAGAGATGAGCAGCA 2057
Qy 1455 TGCACAAAAGCATCTAGAGAAAATTTCAGAGATTCAGAGTTCAGCTGGCAGCA 1514
Db 2058 GCAAGATGAGAGCAGCAGAGATGAGCAGCAGAGATGAGCAGCAGAGATGAGAGCA 2117
Qy 1515 GAGCAGAAAATTGAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1574
Db 2118 GAGCAGCAGAGATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 2177
Qy 1575 CTTACAGCAGAGATGAGAGATCTCAGCAAACTCTTGAAGATTTTCAGAGAAA 1634
Db 2178 GAGATGAGCAGCAGCAGCAGATGAGCAGCAGCAGCAGCAGCAGCAGCAGAGATGAG 2237
Qy 1635 GATCTCAGTACCTGATTAATTAATCAAGTCCAGCAGAACTTTCATAGACTGCTTCT 1694
Db 2238 GCAGCAGCAGCAGATGAGCAGCAGCAGCAGCAGAGAGAGAGAGAGAGAGAGAG 2297
Qy 1695 TACCTCAAAAAGAGCTTGAAGCAAGAGCTGCGCAGCAGCTCCGAGCAGCACT 1754
Db 2298 GCAGAGATGAGAGCAGCAGAGAGCAGAGATGAGAGATGAGAGAGAGATGAGAGAGCA 2357
Qy 1755 GAGCAGGTGAGAGAGAGCAGAGTCAAGCTGAGAGATGAGATGTTTTCACACCA 1814
Db 2358 GAGCAGAGATGAGAGAGCAGAGCAGAGCAGAGATGAGAGAGAGAGAGAGAGATGAGAG 2417
Qy 1815 GCTGAAGAACTGAGAGATATCATGCAACAGCACTCCAGAGCAGAGTCCCTTGA 1874
Db 2418 GCAGAGAGAGATGAGAGCAGCAGCAGAGATGAGAGCAGCAGAGATGAGAGAGT 2477
Qy 1875 GGCAGCGGACCTGAGCAGAAAGCAGAGAGAGAGAGCAGCAGCAGCAGCAGCAG 1934
Db 2478 GAGCAGAGCAGAGATGAGAGAGAGAGAGCAGAGATGAGAGAGAGAGAGAGAGAG 2537
Qy 1935 GGAAGAGCAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1994
Db 2538 GCAAGAGCAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATGAGAGAGT 2597
Qy 1995 GAGAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2054
Db 2598 GGAAGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATGAGAGAGAGAGAGAG 2657

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 4, 2003, 15:03:53 ; Search time 40 Seconds
(without alignments)
4813.383 Million cell updates/sec

Title: US-09-674-237a-3
Perfect score: 6269
Sequence: 1 MAQFPTFGSLDVAIVTVE.....VGLPSPNVYKLTITMDPSQ 1213

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: A.Geneseq.190u03:*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6258.5	99.8	1214	21	AAV57444
2	6258.5	99.8	1715	21	AAV57449
3	5860	93.5	1215	20	AAV32156
4	5847.5	93.3	1220	20	AAV32155
5	5442.5	86.8	1144	20	AAV32154
6	4277	68.2	1031	22	AAU87168
7	3097.5	49.4	1035	22	AAW43519
8	3097.5	49.3	932	22	ABG20575
9	3091.5	49.3	1681	23	AAO17881

10	3070.5	49.0	648	20	AAV32157	Human SH3D1A prote
11	3068	48.9	1197	22	AAW79199	Human protein SEQ
12	3049.5	48.6	1197	21	AAV57445	Mouse Ees2l protei
13	3049.5	48.6	1658	21	AAV57450	Mouse Ees2l protei
14	3046	48.6	1683	21	AAV71160	Rat phosphodiester
15	3034.5	48.4	642	22	AAW93229	Human polypeptide,
16	3012	48.0	641	20	AAV32158	Human SH3D1A prote
17	2812	44.9	595	22	AAW46991	Human protein sequ
18	2273.5	36.3	471	22	AAW93675	Human polypeptide,
19	2058.5	32.8	462	17	AAW05395	Human SH3P17 prote
20	1702	27.1	1097	22	ABW57814	Drosophila melanog
21	1536.5	24.5	936	22	ABG20576	Novel human diagno
22	1388.5	22.1	509	17	AAW05399	Human clone 65 pro
23	1361	21.7	990	22	AAU87256	Novel central nerv
24	1330.5	21.2	530	22	AAU17173	Novel signal trans
25	1183.5	18.9	436	22	AAW93738	Human protein sequ
26	1046	16.7	225	24	ABU71097	Human adipocyte Se
27	816.5	13.0	248	17	AAW05396	Human SH3P18 prote
28	774.5	12.4	182	22	AAO00069	Human polypeptide
29	712	11.4	164	23	ABP41930	Human polypeptide
30	572.5	9.1	233	24	ABU71035	Human adipocyte Se
31	572	9.1	144	22	AAU87484	Novel central nerv
32	572	9.1	144	22	AAW43613	Human polypeptide
33	572	9.1	144	22	AAU19936	Novel human calcit
34	536.5	8.6	864	24	ABU03909	Human expressed pr
35	536.5	8.6	864	24	ABU03910	Human expressed pr
36	536.5	8.6	864	24	ABU03915	Human expressed pr
37	536.5	8.6	864	24	ABU03916	Human expressed pr
38	536.5	8.6	866	24	AAW86194	Nuclear transport
39	536.5	8.6	866	24	ABU03914	Human expressed pr
40	535.5	8.5	864	22	AAW40292	Human polypeptide
41	535.5	8.5	864	24	ABU03912	Human expressed pr
42	527	8.4	813	23	AAO20504	Protein of APP rel
43	503	8.0	754	22	ABW95697	Human protein sequ
44	503	8.0	754	22	ABW95911	Human expressed pr
45	486.5	7.8	896	17	AAW92750	Human EGF receptor

ALIGNMENTS

RESULT 1
ID AAV57444 strand: Protein; 1214 AA.
AAV57444;
AC AAV57444;
XX
XX
DT 28-FEB-2000 (first entry)
XX
XX
DE Mouse Eesl protein sequence.
XX
XX
KW Mouse; murine; Eesl; Ees2; endocytosis; vesicular trafficking;
KW regulation; actin cytoskeleton; detection; cancer; infection;
KW EH-domain and SH3-domain regulator of endocytosis; anticancer;
KW antiproliferative; antiviral.
XX
XX
OS Mus sp.
XX
XX
PN WO9955728-A2.
XX
XX
PD 04-NOV-1999.
XX
XX
PF 27-APR-1999; 99WO-CA00375.
XX
XX
PR 27-APR-1998; 98CA-2230201.
PR 05-FEB-1999; 99US-0118739.
PA (HSCR-) HSC RES & DEV LP.
XX
XX
PI Egan SE, Wang W, Sengar A;
DR WPI; 2000-052802/04.
DR N-PSDB; AA239008, AA239009.

PA (HSCR-) HSC RES & DEV LP.
 XX Egan SE, Wang W, Sengar A;
 XX MPI, 2000-052802/04.
 DR N-PSDB; AA239024, AA239025.
 XX New nucleic acid encoding Eesl and 2 proteins, involved in regulation
 PT of endocytosis, used e.g. for treating cancer or preventing viral
 PT infection -
 PS Claim 14; Page 62-63; 99pp; English.
 XX The present invention specifically describes mammalian Eesl and 2
 CC proteins (1) and their splice variants (Ese = EH-domain and SH3-domain
 CC regulator of endocytosis). (1) are involved in regulation of clathrin-
 CC mediated endocytosis (as a complex with Sepis protein), vesicular
 CC trafficking and actin cytoskeleton. Generally (1) (or its (ant)agonists,
 CC mimetics, fragments and inactive mutants); (1)-specific antibodies (Ab);
 CC sequences antisense to the (1) polynucleotide; agents that downregulate
 CC expression of Ese genes or antagonists of an Ese binding partner are
 CC used to treat diseases associated with undesirable endocytosis and
 CC resulting changes in cellular function. Particularly overexpression of
 CC Eesl is used to block clathrin-mediated endocytosis in vivo or in cell
 CC cultures, while administration of (1) is used to promote endocytosis of
 CC selected cells. (Ant)agonists of (1) or Ab are used to suppress abnormal
 CC proliferation of cells that can be stimulated to proliferate by a growth
 CC factor receptor; and similar compounds (also inactive Ese mutants) can be
 CC used to prevent viral infection. Endocytosis may also be regulated, in
 CC vivo or in cell cultures, by forming an Ese-Sepis complex, then binding
 CC dynamn to the complex. Generally conditions that can be treated include
 CC cancer; abnormal cell division or migration; viral infection; or abnormal
 CC receptor signalling, tissue development or synaptic transmission. The
 CC present sequence represents mouse Eesl protein sequence.
 XX
 SQ Sequence 1715 AA;
 Query Match 99.8%; Score 6258.5; DB 21; Length 1715;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1213; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 481 LEALNDKQHQLGKGLQDIRCRLATQROEIESTNKSRELIATITLQOQLQJESQWGLGRL 540
 DB 481 LEALNDKQHQLGKGLQDIRCRLATQROEIESTNKSRELIATITLQOQLQJESQWGLGRL 540
 QY 541 IFEKQILSDQLKQVQNSLHRDLSLLTLKRALEAKELARQQLDEVERETRSKLOEID 600
 DB 541 IFEKQILSDQLKQVQNSLHRDLSLLTLKRALEAKELARQQLDEVERETRSKLOEID 600
 QY 601 VNNNQKLREIHSQQLQKQSLNAAHLKQKQERKSLILEKQEDNQRQVQERDKQVL 660
 DB 601 VNNNQKLREIHSQQLQKQSLNAAHLKQKQERKSLILEKQEDNQRQVQERDKQVL 660
 QY 661 EHVQOEOPRPKPKHEDELKREDSVRKKEAEERAKPEKQKOSLFFHQBPATKATAQ 720
 DB 661 EHVQOEOPRPKPKHEDELKREDSVRKKEAEERAKPEKQKOSLFFHQBPATKATAQ 720
 QY 721 PMSTTEKGPLTISAQSVKVVYRLLYPESHDEITIQPDIVWVDSQTEBPWLGG 780
 DB 721 PMSTTEKGPLTISAQSVKVVYRLLYPESHDEITIQPDIVWVDSQTEBPWLGG 780
 QY 781 ELKGTGMPFANYAEKIPENEPPTPAKPYTDLTSAAPKLAIRETPAPLPTSSSEPTTP 840
 DB 781 ELKGTGMPFANYAEKIPENEPPTPAKPYTDLTSAAPKLAIRETPAPLPTSSSEPTTP 840
 QY 841 NNWADPSSSTWPSSSNEKPTDNWDMAQPSLTVPSAGQLRORSAFTPATATGSSPSVYL 900
 DB 841 NNWADPSSSTWPSSSNEKPTDNWDMAQPSLTVPSAGQLRORSAFTPATATGSSPSVYL 900
 QY 901 GQGEKVEGLQALYPMWRAKKNHNFNKSVDITVLEQDDMMWFGVQKGMFPKSYVK 960
 DB 901 GQGEKVEGLQALYPMWRAKKNHNFNKSVDITVLEQDDMMWFGVQKGMFPKSYVK 960
 QY 961 LISGPVRSKSTIDTGTPTSPASLKRVAAPAKPAIPGEFFIMYTESSEQDILTFQGD 1020
 DB 961 LISGPVRSKSTIDTGTPTSPASLKRVAAPAKPAIPGEFFIMYTESSEQDILTFQGD 1020
 QY 1021 VIVVT-KKQDGMWGTGVGDKSGVPPSNVYRLKDSGSGAGTKGSLGKKPEIAQVYASYA 1079
 DB 1021 VIVVT-KKQDGMWGTGVGDKSGVPPSNVYRLKDSGSGAGTKGSLGKKPEIAQVYASYA 1080
 QY 1080 ATGPBQLTLAPQQLILIRKKNPGWMEGLQARGKKRQIGWPPANYVKKLSPTSKITPT 1139
 DB 1081 ATGPBQLTLAPQQLILIRKKNPGWMEGLQARGKKRQIGWPPANYVKKLSPTSKITPT 1140
 QY 1140 ELPKTAVQPAVQVGMVDYTAQNDDELAFSKQIINVLNKEDPDMWKEGSGVGLFPS 1199
 DB 1141 ELPKTAVQPAVQVGMVDYTAQNDDELAFSKQIINVLNKEDPDMWKEGSGVGLFPS 1200
 QY 1200 NYVKLTMDPSQ 1213
 DB 1201 NYVKLTMDPSQ 1214

RESULT 3
 AA32156
 ID AAY32156 standard; Protein; 1215 AA.
 XX
 AC AAY32156;
 XX
 DT 01-FEB-2000 (first entry)
 XX
 DE Human SH3D1A protein.
 XX
 KW SH3D1A; human; Down's syndrome; leukemia; cancer;
 KW megakaryocytic abnormality; myeloproliferative disorder;
 KW platelet disorder; neural disorder; thrombocytopenia;
 KW haematopoietic disorder; cognitive dysfunction; microcephaly;
 KW lissencephaly; colpocephaly; holoprosencephaly; diagnosis; therapy.
 XX
 OS Homo sapiens.
 XX
 PN MO953062-A2.
 XX

PD 21-OCT-1999.
 XX 16-APR-1999; 99WO-US08371.
 XX 16-APR-1998; 98US-0082007.
 XX (CEDA-) CEDARS SIGNAL HEALTH SYSTEM.
 PA Korenberg JR, Chen X;
 XX WP1, 1999-633829/54.
 DR N-PSDB; AAZ34572.
 XX Nucleic acid from the human SH3D1A gene and its products, useful for
 PT the diagnosis and treatment of myeloproliferative disorders and
 PT leukaemia
 XX
 PS Claim 14; Fig 11; 99P; English.

XX This sequence represents the protein encoded by the human SH3D1A
 CC cDNA clone 11 (see AAZ34572). SH3D1A contributes to the development
 CC of platelets and the pathogenesis of leukaemias, both in general
 CC and in particular those involving the megakaryocytic lineage. The
 CC gene maps to the small candidate region for low platelets on
 CC chromosome 21. The protein includes SH3 domains and EH domains,
 CC both associated with protein-protein interactions and the latter
 CC with maintenance of the cytoskeleton. At least 3 isoforms of
 CC SH3D1A exist (see AAJ32154-58). The invention provides methods for
 CC the diagnosis and treatment of megakaryocytic abnormality,
 CC myeloproliferative disorder, platelet disorder, acute leukaemia,
 CC neural disorders, thrombocytopenia, platelet disorder on
 CC chromosome 21, low platelets in deletion for 21, association of
 CC gain in chromosome 21 with leukaemias, neural abnormalities,
 CC dysfunctions and disorders including brain malformations and
 CC corresponding cognitive dysfunctions, microcephaly, lissencephaly,
 CC and colpocephaly. Methods are also provided: for suppressing
 CC cells unable to regulate themselves; screening for a somatic
 CC alteration in the SH3D1A gene; monitoring the progress and
 CC adequacy of a treatment; monitoring tumour risk progress or
 CC megakaryocytic abnormality; myeloproliferative disorder,
 CC haematopoietic disorder, platelet disorder or leukaemia; and
 CC screening of drugs for cancer therapy.

XX Sequence 1215 AA;

Query Match 93.5%; Score 5860; DB 20; Length 1215;
 Best Local Similarity 92.9%; Pred. No. 0;
 Matches 1129; Conservative 39; Mismatches 45; Indels 2; Gaps 2;

```

QY 1 MAQPTPGGSLDVAITVEERAKHDQQLSLKPIAGRTTGDQANFFFGSLPQVLAQ 60
DB 1 MAQPTPGGSLDVAITVEERAKHDQQLSLKPIAGRTTGDQANFFFGSLPQVLAQ 60
QY 61 IWLADNMNDGMDQVEFSIAMKLIKLGQYQLPSTLPVWKQOPVAISSAPAGIGI 120
DB 61 IWLADNMNDGMDQVEFSIAMKLIKLGQYQLPSTLPVWKQOPVAISSAPAGIGI 120
QY 121 ASMPPLTAVAPVWMSIPVWGSPLVSSVPPAAPPVPLANGAPVYIQLPAFAHPAATWP 180
DB 121 ASMPPLTAVAPVWMSIPVWGSPLVSSVPPAAPPVPLANGAPVYIQLPAFAHPAATWP 180
QY 181 KSSSFRSGPGSQQLTKYQKASPDVASAPPAEAAVQSSSLTKRQLFNSHDKTMSGHL 240
DB 181 KSSSFRSGPGSQQLTKYQKASPDVASAPPAEAAVQSSSLTKRQLFNSHDKTMSGHL 240
QY 241 TGPQARTILMOSLTPQAOLASITWNLSIDIDQDKLTAEFFILAMHLIDVAMSGQPLPVLP 300
DB 241 TGPQARTILMOSLTPQAOLASITWNLSIDIDQDKLTAEFFILAMHLIDVAMSGQPLPVLP 300
QY 301 PEYIPPSFRVAVSGSGMSVSSSVQDRLPEEPSSSEDEQOP-EKKLPVTFEDKKENFER 359
DB 301 PEYIPPSFRVAVSGSGMSVSSSVQDRLPEEPSSSEDEQOP-EKKLPVTFEDKKENFER 359

```

```

QY 360 GSVLEKRRQALTEQKREORLTAQLERAEOERKERERQEOEAKRQLELEKOLEKORELE 419
DB 361 GNLEKRRQALTEQKREORLTAQLERAEOERKERERQEOEAKRQLELEKOLEKORELE 420
QY 420 ROREERREKIEEREAARELERQOLEWERNRQELLNORNKQOEGSTVILKARKTILEF 479
DB 421 ROREERREKIEEREAARELERQOLEWERNRQELLNORNKQOEDIVLAKAKKTILEF 480
QY 480 ELEALNKKQOLEBKODIRCLATQROEIESTTKSRELRFAETTHQQOLOESQOMLGR 539
DB 481 ELEALNKKQOLEBKODIRCLATQROEIESTTKSRELRFAETTHQQOLOESQOMLGR 540
QY 540 LIPERQILSDOLKQVOONSILHRDSLTLTKALAEKAPQOLREQLDEVERETSXLQEI 599
DB 541 LIPERQILNOLKQVOONSILHRDSLTLTKALAEKAPQOLREQLDEVERETSXLQEI 600
QY 600 DVFNNQLELREIHSKQOLQKRSLEAPRLKQKEERKSLELEKQEDAQORVQERDKQW 659
DB 601 DIFNNQLELREIHNKQOLQKQSMEAERLKQKEERKILELEKQKEAQRRQERDKQW 660
QY 660 LEHQOE-EOPRPKPHEDRLKREDSVRKKEAEERAKPEWQDKSRLFHPHOBPAKLAT 718
DB 661 LEHQOEDHQRPKPHHEEKLKREESVKKKQGEKKQEAQDKGLRFLHQHOPAPKAV 720
QY 719 QAPSTTEKGPLTISAQESVYVYVYRALYPFESRSHDEITIQPDIVWVDESQTEGPGL 778
DB 721 QAPSTAEKGPLTISAQENKVVYVYRALYPFESRSHDEITIQPDIVWVDESQTEGPGL 780
QY 779 GEGELKGTGWFPANYAEKIPENEVTPAKVTDITSAAPAPKLARETPAPLAVTSSEST 838
DB 781 GEGELKGTGWFPANYAEKIPENEVPAPEVTPDSTAPAPKLARETPAPLAVTSSEST 840
QY 839 TPNWMAFPSSSWPSSSNKEPBTDMWMAAOPSLTVSSAQGLRQSAFTPATATGSSSP 898
DB 841 TPNWMAFPSSSWPSTNEKEPBTDMWMAAOPSLTVSSAQGLRQSAFTPATATGSSSP 900
QY 899 VLGGKEKEGLOAALTPWRAKDNHLPFNKSDVITYLEQDDMMFGEVQOKGMPKSY 958
DB 901 VLGGKEKEGLOAALTPWRAKDNHLPFNKSDVITYLEQDDMMFGEVQOKGMPKSY 960
QY 959 VKLISGPRKSTSIDTGPTESSPASLKEVASPAKPAIPGEEFIAMTYESESQDLTFQ 1018
DB 961 VKLISGPRKSTSIDTGPTESSPASLKEVASPAKPAIPGEEFIAMTYESESQDLTFQ 1020
QY 1019 GDVIVYTKKQDMMTGVGDKSGVFPNNYVRLKDSBESGTAAGTSGKKEINQVASY 1078
DB 1021 GDVIVYTKKQDMMTGVGDKSGVFPNNYVRLKDSBESGTAAGTSGKKEINQVASY 1080
QY 1079 AATGPEDLTAPGQLILIRKKNPGMWEGELQARGKKRQIGWFPANYVKLSPESTKITP 1138
DB 1081 AATGPEDLTAPGQLILIRKKNPGMWEGELQARGKKRQIGWFPANYVKLSPESTKITP 1140
QY 1139 TELPKTAVQAPVCQVIGMYDTAQNDELAEKSKQIINVLNKEDPDMWKGESQVGLFP 1198
DB 1141 TELPKTALAAVCVIGMYDTAQNDELAEKSKQIINVLNKEDPDMWKGESQVGLFP 1200
QY 1199 SNVYKLTITDMDPSQ 1213
DB 1201 SNVYKLTITDMDPSQ 1215

```

RESULT 4

AAJ32155
 ID AAJ32155 standard; Protein; 1220 AA.

XX AAJ32155;

XX 01-FEB-2000 (first entry)

XX Human SH3D1A protein.

XX SH3D1A; human; Down's syndrome; leukaemia; cancer;
 KW megakaryocytic abnormality; myeloproliferative disorder;

KM platelet disorder; neural disorder; thrombocytopenia;
 KM haematopoietic disorder; cognitive dysfunction; microcephaly;
 KM lissencephaly; colpocephaly; holoprosencephaly; diagnosis; therapy:
 OS Homo sapiens.
 XX MO9953062-A2.
 XX
 XX 21-OCT-1999.
 PD
 XX 16-APR-1999; 99MO-US08371.
 PF
 XX 16-APR-1998; 98US-0082007.
 PR
 XX (CEDA-) CEDARS SINAI HEALTH SYSTEM.
 PA
 XX korenberg JR, Chen X;
 PI
 XX MPI; 1999-633829/54.
 DR
 XX N-PSDB; AA234571.
 XX
 XX Nucleic acid from the human SH3D1A gene and its products, useful for
 PT the diagnosis and treatment of myeloproliferative disorders and
 PT leukaemia -
 PS
 PS Claim 14; Fig 9; 99pp; English.
 XX
 CC This sequence represents the protein encoded by the human SH3D1A
 CC CDNA clone 21 (see AA234571). SH3D1A contributes to the development
 CC of platelets and the pathogenesis of leukemias, both in general
 CC and in particular those involving the megakaryocytic lineage. The
 CC gene maps to the small candidate region for low platelets on
 CC chromosome 21. The protein includes SH3 domains and EH domains,
 CC both associated with protein-protein interactions and the latter
 CC with maintenance of the cytoskeleton. At least 3 isoforms of
 CC SH3D1A exist (see AA232154-58). The invention provides methods for
 CC the diagnosis and treatment of megakaryocytic abnormality,
 CC myeloproliferative disorder, platelet disorder, acute leukaemia,
 CC neural disorders, thrombocytopenia, platelet disorder on
 CC chromosome 21, low platelets in deletion for 21, association of
 CC gains in chromosome 21 with leukemias, neural abnormalities,
 CC dysfunctions and disorders including brain malformations and
 CC corresponding cognitive dysfunctions, microcephaly, lissencephaly,
 CC and colpocephaly. Methods are also provided for suppressing
 CC cells unable to regulate themselves, screening for a somatic
 CC alteration in the SH3D1A gene; monitoring the progress and
 CC adequacy of a treatment; monitoring tumour risk progress or
 CC megakaryocytic abnormality, myeloproliferative disorder,
 CC haematopoietic disorder, platelet disorder or leukaemia; and
 CC screening of drugs for cancer therapy.
 XX
 XX Sequence 1220 AA;
 SQ
 Query Match 93.3%; Score 5847.5; DB 20; Length 1220;
 Best local Similarity 92.5%; Pred. No. 0;
 Matches 1199; Conservative 39; Mismatches 45; Indels 7; Gaps 3;

QY 241 TGPARTIMOSSLPQALASIMNLSIDIPQDGLTAEEFIAMHLIDVAMSGPLPVLP 300
 |||||
 DB 241 TGPARTIMOSSLPQALASIMNLSIDIPQDGLTAEEFIAMHLIDVAMSGPLPVLP 300
 QY 301 PEYIPPSFRVRVSGSGMSVISSSVDORLPEEPPSSDEEQP-EKKLPVTFEDKRENFER 359
 |||||
 DB 301 PEYIPPSFRVRVSGSGMSVISSSVDORLPEEPPSSDEEQP-EKKLPVTFEDKRENFER 359
 QY 360 GSVELEKRRQALLQQRKEQERLAQLERAEQERKEERQOEAKROLLEKOLEKORELE 419
 |||||
 DB 360 GSVELEKRRQALLQQRKEQERLAQLERAEQERKEERQOEAKROLLEKOLEKORELE 419
 QY 420 RQREERREKREIRREBAKKELEERORQLEERRRROELNQRKKEOGTVLAKRRITLFE 479
 |||||
 DB 420 RQREERREKREIRREBAKKELEERORQLEERRRROELNQRKKEOGTVLAKRRITLFE 479
 QY 481 ELEALNDKKHOLEGKLODIRCLATQROEIESTNKSRELRIAEITHLOQLOESQOMLGR 540
 |||||
 DB 481 ELEALNDKKHOLEGKLODIRCLATQROEIESTNKSRELRIAEITHLOQLOESQOMLGR 540
 QY 540 LIPEKOILSDOLKOVQONSILHRDLSLTLKRALEAKELAQQOLREQLDEVERETRSKLOEI 599
 |||||
 DB 540 LIPEKOILSDOLKOVQONSILHRDLSLTLKRALEAKELAQQOLREQLDEVERETRSKLOEI 599
 QY 600 DVENNOLKELREITHSKOOLQOKORSLPEARLKQKEORKSLELEKOKEDARQROERDKOM 659
 |||||
 DB 600 DVENNOLKELREITHSKOOLQOKORSLPEARLKQKEORKSLELEKOKEDARQROERDKOM 659
 QY 660 LEHVQOE-EQPPRRKHEEDRLKREDSYRKKEAEERAKPEMOKOSRLPHPHQEPKALAT 718
 |||||
 DB 660 LEHVQOE-EQPPRRKHEEDRLKREDSYRKKEAEERAKPEMOKOSRLPHPHQEPKALAT 718
 QY 719 QAPWSTTEGRLTISAQESVKKVYRYRALYPPSSRSDEITITOPGDIW-----VDESQTC 773
 |||||
 DB 719 QAPWSTTEGRLTISAQESVKKVYRYRALYPPSSRSDEITITOPGDIW-----VDESQTC 773
 QY 774 EBGWLGELGKGTGMPANYAKTIPENEVPTPAKPTDLSAPAPGLARETPAPLPYVS 833
 |||||
 DB 774 EBGWLGELGKGTGMPANYAKTIPENEVPTPAKPTDLSAPAPGLARETPAPLPYVS 833
 QY 834 SEBSTTPNNMADPSSSTWSPSSNEKETDMDTMAQPSLTVPSAGQLRORSAFTPATATG 893
 |||||
 DB 834 SEBSTTPNNMADPSSSTWSPSSNEKETDMDTMAQPSLTVPSAGQLRORSAFTPATATG 893
 QY 894 SSBSPVLGGEGVEGLQALATYPRACKDNHLNFKSDVITYLEQDDMMFGEVQCKGM 953
 |||||
 DB 894 SSBSPVLGGEGVEGLQALATYPRACKDNHLNFKSDVITYLEQDDMMFGEVQCKGM 953
 QY 954 PFKSVYKLSGPRKSTSIDGPTSPSLKRVASPAKPAIPIGEFFIMYTESSEQGD 1013
 |||||
 DB 954 PFKSVYKLSGPRKSTSIDGPTSPSLKRVASPAKPAIPIGEFFIMYTESSEQGD 1013
 QY 1014 LTFQOGDVIIVTKGDMMTGVGDKSGVFPNNYVRLKDSGSGTAGTSGLKKPEIAQ 1073
 |||||
 DB 1014 LTFQOGDVIIVTKGDMMTGVGDKSGVFPNNYVRLKDSGSGTAGTSGLKKPEIAQ 1073
 QY 1074 VIASITATPEQDITLAPGQLIIRKKNPGCWMEGELQARGKKRQIGWFPANYVKLISPGT 1133
 |||||
 DB 1074 VIASITATPEQDITLAPGQLIIRKKNPGCWMEGELQARGKKRQIGWFPANYVKLISPGT 1133
 QY 1134 SKITTEPEKSTALAAVCVIGMYDTAQNDELAFSKQOILNVLKPEDPMWKGVSQO 1193
 |||||
 DB 1134 SKITTEPEKSTALAAVCVIGMYDTAQNDELAFSKQOILNVLKPEDPMWKGVSQO 1193
 QY 1194 VGLFPSNYYKLTMDMPSSQO 1213
 |||||
 DB 1194 VGLFPSNYYKLTMDMPSSQO 1213
 QY 1201 VGLFPSNYYKLTMDMPSSQO 1220
 |||||
 DB 1201 VGLFPSNYYKLTMDMPSSQO 1220
 RESULT 5
 AA32154
 ID AA32154 standard; Protein; 1144 AA.
 XX

AC	AA32154;	
XX		
DT	01-FEB-2000	(first entry)
XX		
DE	Human SH3D1A protein.	
XX		
KW	SH3D1A; human; Down's syndrome; leukaemia; cancer;	
KW	megakaryocytic abnormality; myeloproliferative disorder;	
KW	platelet disorder; neural disorder; thrombocytopenia;	
KW	haematopoietic disorder; cognitive dysfunction; microcephaly;	
XX	lissencephaly; colpocephaly; holoprosencephaly; diagnosis; therapy.	
OS	homo sapiens.	
XX		
PH	Key.	Location/Qualifiers
FT	Domain	15..102
FT		/note= "EH1 domain"
FT	Domain	215..310
FT		/note= "EH2 domain"
FT	Domain	740..800
FT		/note= "SH3-1 domain"
FT	Domain	908..966
FT		/note= "SH3-2 domain"
FT	Domain	999..1062
FT		/note= "SH3-3 domain"
FT	Domain	1080..1138
FT		/note= "SH3-4 domain"
XX		
PN	W09953062-A2.	
PD	21-OCT-1999.	
XX		
PP	16-APR-1999;	99WO-US08371.
XX		
PR	16-APR-1998;	98US-0082007.
XX		
PA	(CEDA-) CEDARS SINAI HEALTH SYSTEM.	
XX		
PI	Korenberg JR, Chen X;	
XX		
DR	WPI, 1999-633829/54.	
DR	N-PSDB; AA234570.	
XX		
PT	Nucleic acid from the human SH3D1A gene and its products, useful for	
PT	the diagnosis and treatment of myeloproliferative disorders and	
PT	leukaemia	
XX	-	
PS	Claim 14; Fig 5; 99p; English.	
XX		
CC	This sequence represents the protein encoded by the novel human	
CC	SH3D1A gene (see AA234570). SH3D1A contributes to the development	
CC	of platelets and the pathogenesis of leukaemias, both in general	
CC	and in particular those involving the megakaryocytic lineage. The	
CC	gene maps to the small candidate region for low platelets on	
CC	chromosome 21. The protein includes SH3 domains and EH domains,	
CC	both associated with protein-protein interactions and the latter	
CC	with maintenance of the cytoskeleton. At least 3 isoforms of	
CC	SH3D1A exist (see AA32154-58). The invention provides methods for	
CC	the diagnosis and treatment of megakaryocytic abnormality,	
CC	myeloproliferative disorder, platelet disorder, acute leukaemia,	
CC	neural disorders, thrombocytopenia, platelet disorder on	
CC	chromosome 21, low platelets in deletion for 21, association of	
CC	gains in chromosome 21 with leukaemias, neural abnormalities,	
CC	dysfunctions and disorders including brain malformations and	
CC	corresponding cognitive dysfunctions, microcephaly, lissencephaly,	
CC	and colpocephaly. Methods are also provided: for suppressing	
CC	cells unable to regulate themselves; screening for a somatic	
CC	alteration in the SH3D1A gene; monitoring the progress and	
CC	adequacy of a treatment; monitoring tumour risk progress or	
CC	megakaryocytic abnormality, myeloproliferative disorder,	
CC	haematopoietic disorder, platelet disorder or leukaemia; and	
CC	screening of drugs for cancer therapy.	

Seq	Sequence	1144 AA	96.8%;	Score 5442.5;	DB 20;	Length 1144;
Query Match	87.2%;	Pred. No. 0;				
Best Local Similarity	Matches 1060;	Conservative 37;	Mismatches 45;	Indels 73;	Gaps 3;	
QY	1 MAQFPPTPGSLDVAITVEERAKHDQFSLKPIAGITTDQAKRNFPPFQSGLPQVLAQ	60				
DB	1 MAQFPPTPGSLDVAITVEERAKHDQFSLKPIAGITTDQAKRNFPPFQSGLPQVLAQ	60				
QY	61 IMAIDAMNDMDRMQVSESIAMKLIKTLQGYQLPSTLPVYKQOPVAISSAPAFGIGI	120				
DB	61 IMAIDAMNDMDRMQVSESIAMKLIKTLQGYQLPSTLPVYKQOPVAISSAPAFGIGI	120				
QY	121 ASMPPLTAVAPVWGMSIPVVGMSPTLVSSVTPAAPPPLANGAPVIOPLPAFAHPATLP	180				
DB	121 ASMPPLTAVAPVWGMSIPVVGMSPTLVSSVTPAAPPPLANGAPVIOPLPAFAHPATLP	180				
QY	181 KSSSFSSRGSSQSLNTKIQKQSPDVASAPAAAMAPQSSRLKYROLFNSHDTMSGHL	240				
DB	181 KSSSFSSRGSSQSLNTKIQKQSPDVASAPAAAMAPQSSRLKYROLFNSHDTMSGHL	240				
QY	241 TGPARITLMSSIPQAKLASINMLSDIDQDKLTAEFIAMLLIVAMSGCLPVLPLP	300				
DB	241 TGPARITLMSSIPQAKLASINMLSDIDQDKLTAEFIAMLLIVAMSGCLPVLPLP	300				
QY	301 PEYIPSPFRVRSSGMSVSISSSVDRLPPEPSSSEDOQP-EKKLPVTFDKKRENER	359				
DB	301 PEYIPSPFRVRSSGMSVSISSSVDRLPPEPSSSEDOQP-EKKLPVTFDKKRENER	359				
QY	360 GSVELEKRRALLQOKKEQERLAQLERABEERKEREROEBAKROLELEKOLEKORLE	419				
DB	361 GNLBEKRRQALLQOKKEQERLAQLERABEERKEREROEBAKROLELEKOLEKORLE	420				
QY	420 ROREBERKELEERBAKRELEROROLEMENRKOELLNOKNKOEGVVLVAKARKTLEF	479				
DB	421 ROREBERKELEERBAKRELEROROLEMENRKOELLNOKNKOEDVLVAKAKKTLEF	480				
QY	480 ELEALNDKQOLEGKLODIRCLATOROLEBISTNKSRELIATHTHLOOLOESQOMLGR	539				
DB	481 ELEALNDKQOLEGKLODIRCLATOROLEBISTNKSRELIATHTHLOOLOESQOMLGR	540				
QY	540 LIPEKQIISDOLKOVQONSILHRDSLLTKALAEKELARQOLREOLDEVEFTSKLOEI	599				
DB	541 LIPEKQIISDOLKOVQONSILHRDSLLTKALAEKELARQOLREOLDEVEFTSKLOEI	600				
QY	600 DVPNNOLKEALEHISKQOLKORSLLEAKLEKOKOERKSLLEKOKEDQAKORVOERDKQW	659				
DB	601 DFPNNOLKEALEHINKQOLKOKSWEABRLKQKOEKKIILEKQKBEAQKRAQERDKQW	660				
QY	660 LEHVQOE-EQRPKPKPHEDELKEDSRKKAEERAKPEWODKOSRLFHPHOEPAKLAT	718				
DB	661 LEHVQOEDEHQRPKLHBEKLLKKEESVYKKKDGSEKQGEAKDGLGRLFHQHGRPAKAV	720				
QY	719 QAPWSTTEKGLTISAQESVKVYVYRALYPPESRSHDEITIQPGDIYVWDESGTGEPMWL	778				
DB	721 QAPWSTAEKGLTISAQENVKVYVYRALYPPESRSHDEITIQPGDIYVWDESGTGEPMWL	780				
QY	779 GGELKKGKGMFPANYAEKIPENEVPTAKVPTDLSAPAKLAREHPAPALPYVSSSEST	838				
DB	781 GGELKKGKGMFPANYAEKIPENEVPAKVPJTDSAPAKLAREHPAPALVYSSSEST	840				
QY	839 TPNNMADPSSSPSSSEKPEETDMDMDTMAAOPSLTVSAGQLRORSAFTPATAGSSSPF	898				
DB	841 TPNNMADPSSSPVSTNEKEPETDMDMDMAAOPSLTVSAGQLRORSAPFTPATAGSSSP	900				
QY	899 VLAGGEKEVGLQAQALVYPRAKKDNHLNFKNSDIVITVLEQDDMMWFGEVQOGKGMFPKSY	958				
DB	901 VLAGGEKEVGLQAQALVYPRAKKDNHLNFKNSDIVITVLEQDDMMWFGEVQOGKGMFPKSY	960				
QY	959 VKLISGPRKSTSIDTGTBSPALAKVAPSAAPALPGESEFIAMTYTESSEQDGLTQQ	1018				
DB	961 VKLISGPRKSTSIDTGTBSPALAKVAPSAAPALPGESEFIAMTYTESSEQDGLTQQ	1000				

QY 1019 GVIVVTKKDGMMVTGVGDKSGVFPNSNYRLKDSSESGTAGKTSGLKKPEIAQVIASY 1078
Db 1001 -----EIAQVIASY 1009
QY 1079 AATGPEQLTLAPQGLILIRKNFGWMEGELQARGKKRQIGWFPANVYKLLSPGTSKITP 1138
Db 1010 TATGPEQLTLAPQGLILIRKNFGWMEGELQARGKKRQIGWFPANVYKLLSPGTSKITP 1069
QY 1139 TELPKTAVGPAVQGVGMVDYTAQNDELAFSGQIINLVNKEDPMMKGEVSGQVGLTP 1198
Db 1070 TEPKSTALAAVCQVIGMVDYTAQNDELAFNKGQIINLVNKEDPMMKGEVNGQVGLTP 1129
QY 1199 SNVYKLLTMDPSQ 1213
Db 1130 SNVYKLLTMDPSQ 1144

RESULT 6
AAU87168
ID AAU87168 standard; Protein; 1031 AA.
XX
AC AAU87168;
DT 05-JUN-2002 (first entry)
XX
DE Novel central nervous system protein #78.
XX
KW Central nervous system; CNS; autoimmune disease; rheumatoid arthritis;
KW hyperproliferative disorder; neoplasm; cardiovascular disorder;
KW cardiac arrest; cerebrovascular disorder; ischaemia; angiogenesis;
KW nervous system disorder; Alzheimer's disease; AIDS; ocular disorder;
KW acquired immunodeficiency virus; dysphagia; gastrointestinal disorder;
KW adenocarcinoma; reproductive system disorder; testicular feminisation;
KW endocrine disorder; diabetes; cancer; leukaemia; neovascularisation;
KW respiratory disorder; renal disorder; kidney failure; blood disorder;
KW myocardial infarction; wound healing; cell proliferation; skin aging;
KW food additive; food preservative; gene therapy.
XX
OS Homo sapiens.
XX
PN WO200155318-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01332.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-018674.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 11-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 14-JUL-2000; 2000US-0217496.
PR 26-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.

PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 23-AUG-2000; 2000US-0227182.
PR 30-AUG-2000; 2000US-0227009.
PR 01-SEP-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229349.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0232403.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239335.
PR 13-OCT-2000; 2000US-0239337.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.

PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Barash SC, Ruben SM;

XX WPI; 2001-581633/65.

DR N-PSDB; ABK43498.

PT New isolated nucleic acid encoding a protein for diagnosing,
 PT preventing, treating or ameliorating medical conditions and used as
 PT food additives or preservatives -

PS Claim 9; SEQ ID No 686; 837bp; English.

XX The invention describes an isolated nucleic acid molecule (I) encoding a
 CC novel central nervous system protein. (I) and polypeptides (III) encoded
 CC by (I), are used to treat a medical conditions and in diagnosis of a
 CC pathological condition. Disorders which are diagnosed or treated include
 CC autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
 CC disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
 CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
 CC angiogenesis, nervous system disorders e.g. Alzheimer's disease and
 CC amyotrophic lateral sclerosis, infections caused by bacteria, viruses
 CC e.g. Acquired immunodeficiency virus (AIDS) and fungi, ocular disorders
 CC e.g. corneal infection, gastrointestinal disorders e.g. dysphagia,
 CC adenocarcinomas and irritable bowel syndrome, reproductive system
 CC disorders e.g. testicular feminisation, endocrine disorders e.g. diabetes
 CC and pituitary dwarfism, cancers and disorders at the cellular level e.g.
 CC leukaemia, disorders involving neovascularisation e.g. melaninocytes,
 CC respiratory disorders e.g. nonallergic rhinitis, renal disorders e.g.
 CC acute kidney failure and blood related disorders e.g. myocardial
 CC infarction. The polypeptides can also be used to aid wound healing and
 CC epithelial cell proliferation, to prevent skin aging due to sunburn, to
 CC maintain organs before transplantation, for supporting cell culture of
 CC primary tissues, to regenerate tissues and in chemotaxis. The
 CC polypeptides can also be used as a food additive or preservative to

CC increase or decrease storage capabilities, fat content, lipid, protein,

Query Match 68.2%; Score 4277; DB 22; Length 1031;
 Best Local Similarity 92.3%; Pred. No. 3.7e-266;
 Matches 838; Conservative 28; Mismatches 36; Indels 6; Gaps 3;

QY 1 MAQPTPFGSLDVMATVEERAKHQOFLSLKPIAGFTIGDQARFFPQSGLPQVLAQ 60
 DB 8 MAQPTPFGSLDVMATVEERAKHQOFLSLKPIAGFTIGDQARFFPQSGLPQVLAQ 67
 QY 61 IMLADNMNDGRMDYVEFSIAMKLITKLQGYLPSTLPVWKQOVPVASSAPAFGIGI 120
 DB 68 IMLADNMNDGRMDYVEFSIAMKLITKLQGYLPSTLPVWKQOVPVASSAPAFGIGI 127
 QY 121 ASMPPLTAAPVPMGSIPIVVGMSPPVSSVPAAVPLANGAPVPOPLPAFAHPAATWP 180
 DB 128 ASMPPLTAAPVPMGSIPIVVGMSPPVSSVPAAVPLANGAPVPOPLPAFAHPAATWP 187
 QY 181 KSSSFRSGSGSLNTKLQKQSPDVASAPPAEMAVPOSSRLKYRQLFNSHDKTMSGHL 240
 DB 188 KSSSFRSGSGSLNTKLQKQSPDVASAPPAEMAVPOSSRLKYRQLFNSHDKTMSGHL 247
 QY 241 T---GPOARTIMQSSLPQALASIMNLSIDIDQGLTAEFTLMLHLDIVAMSGQLP 296
 DB 248 TGSCIGPQARTIMQSSLPQALASIMNLSIDIDQGLTAEFTLMLHLDIVAMSGQLP 307
 QY 297 PVLPEYIPPSFRFVSSGMSVSISSSVQDLPEPSSSDEQEP-EKLPVPEKKRE 355
 DB 308 PVLPEYIPPSFRFVSSGMSVSISSSVQDLPEPSSSDEQEP-EKLPVPEKKRE 367
 QY 356 NFRSGVLEKRRQALLQKQERLAQERAEORKEERQOEAKROLJLEKOLEKQ 415
 DB 368 NFRSGVLEKRRQALLQKQERLAQERAEORKEERQOEAKROLJLEKOLEKQ 427
 QY 416 RELERQEEERKEIERREAKLEERQOLEMERNRQGLNQRNKEGEGTVLXARK 475
 DB 428 RELERQEEERKEIERREAKLEERQOLEMERNRQGLNQRNKEGEGTVLXARK 487
 QY 476 TLEFELALNDKKGHLEGLQDIRCRLATROEIESTNKRELRIRAEITLHQOQLQESSQ 535
 DB 488 TLEFELALNDKKGHLEGLQDIRCRLATROEIESTNKRELRIRAEITLHQOQLQESSQ 547
 QY 536 MLGRILPEKQIISDQLKQVOQNSLHRDLSLTLLKRALAEKELARQOLREQLDEVERETRSK 595
 DB 548 MLGRILPEKQIISDQLKQVOQNSLHRDLSLTLLKRALAEKELARQOLREQLDEVERETRSK 607
 QY 596 LOEIDVENNQLKEIRLHSHKQOLQOKRSLERARLKQKEGRKSLLEKQEDAQRRVQER 655
 DB 608 LOEIDVENNQLKEIRLHSHKQOLQOKRSLERARLKQKEGRKSLLEKQEDAQRRVQER 667
 QY 656 DKQMLEHVQOE-EOPRRKHEEDRLKREDSYKKEAEERAKEMQOKORLFFHQEPA 714
 DB 668 DKQMLEHVQOE-EOPRRKHEEDRLKREDSYKKEAEERAKEMQOKORLFFHQEPA 727
 QY 715 KLAQAWSTTEKGPLTISAQESKYVYRYALYPFESRSHDEITIGDITWVDESQTGE 774
 DB 728 KPAVQAWSTTEKGPLTISAQESKYVYRYALYPFESRSHDEITIGDITWVDESQTGE 787
 QY 775 PGWLGELKQKTGFPANVAEKIPENEVPTPAKPVTDLTSAAPKALRETAPAVTSS 834
 DB 788 PGWLGELKQKTGFPANVAEKIPENEVPTPAKPVTDLTSAAPKALRETAPAVTSS 847
 QY 835 EPSTTPNNMADFSTWSSSNEKETNNWTMAAQPSLITPSAQQLRQSAFPTATGTS 894
 DB 848 EPSTTPNNMADFSTWSSSNEKETNNWTMAAQPSLITPSAQQLRQSAFPTATGTS 907
 QY 895 SPSPVLQ 902
 DB 908 SPSPVLQ 915

RESULT 7
 AAM43519

ID AAM43519 standard; Protein; 1035 AA.
XX
AC AAM43519;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 197.
XX
XX Human; antiaerthetic; antirheumatic; antiproliferative; vasotropic;
KW cerebroprotective; nootropic; neuroprotective; antibacterial; virucide;
KW fungicide; ophthalmological; cytostatic; immunosuppressive; nootropic;
KW neuroprotective; antiallergic; hepatotropic; antidiabetic;
KW antiinflammatory; antituber; vulnerary; anticonvulsant; antibacterial;
KW antiparasitic; cardiac; gene therapy; cancer; immune disorder;
KW cardiovascular disorder; neurological disease; infection; human.
XX
OS Homo sapiens.
XX
XX WO200155308-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01309.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189674.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 14-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225477.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.

PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0246207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.

PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX
 XX Rosen CA, Barash SC, Ruben SW;
 XX
 DR WPI, 2001-488781/53.
 DR N-PSDB; AAI63825.
 XX
 PT New isolated nucleic acids and polypeptides, useful for diagnosing,
 PT treating and/or preventing human diseases and disorders -
 XX
 PS Claim 11; SEQ ID NO 197; 664pp + Sequence listing; English.
 XX
 CC The invention relates to human polynucleotides (AAI63803-AAI64012) and
 CC the encoded proteins (AAM434497-AAM43660) useful for preventing, treating
 CC or ameliorating medical conditions e.g. by protein or gene therapy. The
 CC genes were isolated from a range of human tissues disclosed in the
 CC specification. The nucleic acids, proteins, antibodies and (ant)agonists
 CC are useful in the diagnosis, treatment and prevention of: (a) cancer,
 CC e.g. breast and ovarian cancer and other cancers of the adrenal gland,
 CC bone, bone marrow, breast, gastrointestinal tract, liver, lung, or
 CC urogenital; (b) immune disorders e.g. Addison's disease, allergies,
 CC autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus,
 CC Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
 CC and parasitic infections.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX
 XX Sequence 1035 AA;

Query Match 68.2% Score 4277; DB 22; Length 1035;
 Best Local Similarity 92.3%; Pred No. 3.8e-266;
 Matches 838; Conservative 28; Mismatches 36; Indels 6; Gaps 3;

QY 1 MAQPTPTPGSLDVAIVTEERAKHDQFLSLKPIAGFTGDAANFFQSGLPQPVLAQ 60
 DB 8 MAQPTPTPGSLDVAIVTEERAKHDQFLSLKPIAGFTGDAANFFQSGLPQPVLAQ 67
 QY 61 IVALADNNNDGMDQVFSIANKLIKIKIQCQYQLPSTLPVVKQOPVAISSAPAGIGCI 120
 DB 68 IVALADNNNDGMDQVFSIANKLIKIKIQCQYQLPSTLPVVKQOPVAISSAPAGIGCI 127
 QY 121 ASMPPLTAVAPVPMGSIPIVGVSPPLVSSVPPAAVPLANGAPVPIOTPLPAFHPAATMP 180
 DB 128 ASMPPLTAVAPVPMGSIPIVGVSPPLVSSVPPAAVPLANGAPVPIOTPLPAFHPAATMP 187
 QY 181 KSSSFSGSGPGLTKLQKASQFVVASAPAAEAVVQSSRLKRYOLFNSHDKTMSGHL 240
 DB 188 KSSSFSGSGPGLTKLQKASQFVVASAPAAEAVVQSSRLKRYOLFNSHDKTMSGHL 247
 QY 241 T---GPOARTLMOSSLPOAQLASIMWISDIDQGLTAEEFIILAMHLDIVAMGQPLP 296
 DB 248 TGSCGPGPOARTLMOSSLPOAQLASIMWISDIDQGLTAEEFIILAMHLDIVAMGQPLP 307
 QY 297 PVLPEYIIPPSFRVRSSGSGSVISSSVQGLPEPSESEBQGP-EKKLPTFEDKKRE 355

DB 308 PVLPEYIIPPSFRVRSSGSGSVISSSVQGLPEPSESEBQGP-EKKLPTFEDKKRE 367
 QY 356 NFERGSAVELEKRRQALLIQKQKEOERLAQLERAEQEKERBRQEQAKQLELEKQLEKQ 415
 DB 368 NFERGSAVELEKRRQALLIQKQKEOERLAQLERAEQEKERBRQEQAKQLELEKQLEKQ 427
 QY 416 RELEROREBERREKKEIERRBAKRELEROROLEMERBRNROELNORNEQEGTVVLKARK 475
 DB 428 RELEROREBERREKKEIERRBAKRELEROROLEMERBRNROELNORNEQEGTVVLKARK 487
 QY 476 TLFELBALNDKKHLEGKQDIDRCRLATQROEIESTNKSRELRIAEITHLQOQLQESQ 535
 DB 488 TLFELBALNDKKHLEGKQDIDRCRLATQROEIESTNKSRELRIAEITHLQOQLQESQ 547
 QY 536 MLGRLIPEKQILSLQKQVQONSILHRSLSLTALKALPAKELARQALREQLDEVRETRSK 595
 DB 548 MLGRLIPEKQILSLQKQVQONSILHRSLSLTALKALPAKELARQALREQLDEVRETRSK 607
 QY 596 LQETIDVNNQKELREIHSKQQLKORSLEAARLKQKQEKRSKLELEKQEKDAQRVQER 655
 DB 608 LQETIDVNNQKELREIHSKQQLKORSLEAARLKQKQEKRSKLELEKQEKDAQRVQER 667
 QY 656 DKQMLEHVQOE-EQPRPRKPHEDRLKREDSVRKKEAEERAKPEMODKOSRLFFPHQEP 714
 DB 668 DKQMLEHVQOE-EQPRPRKPHEDRLKREDSVRKKEAEERAKPEMODKOSRLFFPHQEP 727
 QY 715 KLAQAPWSTTEKGLPLTISAQESVYVYRYALYFBSRSHDEITIQGDIYMWDESQGE 774
 DB 728 KPAQVQAWSTTEKGLPLTISAQESVYVYRYALYFBSRSHDEITIQGDIYMWDESQGE 787
 QY 775 PGWIGELKGTGTFPANYAEKIPENEVPPAPKAVTDLTSAAPAKLALRETPAPLPYTS 834
 DB 788 PGWIGELKGTGTFPANYAEKIPENEVPPAPKAVTDLTSAAPAKLALRETPAPLPYTS 847
 QY 835 EPSTTPNNWADFSSTWSSSNEKEPDTNMDTMAAPSLTYPSAQQLRQSAFTPATATGS 894
 DB 848 EPSTTPNNWADFSSTWSSSNEKEPDTNMDTMAAPSLTYPSAQQLRQSAFTPATATGS 907
 QY 895 SPSPVLAQ 902
 DB 908 SPSPVLAQ 915

RESULT 8
 ABG20575
 ID ABG20575 standard; Protein; 932 AA.
 AC ABG20575;
 DT 13-FEB-2002 (first entry)
 XX
 XX DE Novel human diagnostic protein #20566.
 XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 XX KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX OS Homo sapiens.
 XX PN MO200175067-A2.
 XX PD 11-OCT-2001.
 XX PF 30-MAR-2001; 2001WO-US08631.
 XX PR 31-MAR-2000; 2000US-0540217.
 XX PR 23-AUG-2000; 2000US-0649167.
 XX PA (HYSE-) HYSEQ INC.
 XX PI Drmanac RT, Liu C, Tang YT;
 XX WPI, 2001-639362/73.

DR N-PSDB; AAS84762.
 XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity
 XX
 PS Claim 20; SEQ ID No 50934; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 932 AA;
 Query Match 49.4%; Score 3097.5; DB 22; Length 932;
 Best Local Similarity 80.0%; Pred. No. 2.3e-190;
 Matches 603; Conservative 26; Mismatches 26; Indels 99; Gaps 3;
 QY 461 NNEOEGTVLKKARKTLEFELEALNDKKQLQESKQLODRCRLATQOELESTNKSSEAI 520
 DB 2 NNEOEDIVLVKAKKTLFELEALNDKKQLQESKQLODRCRLATQOELESTNKSSEAI 61
 QY 521 AEITHLQOOLQESQOQMLGRLIPEKQILSDQLKQVQNSLHRSLSLTALKALEKELARQO 580
 DB 62 AEITHLQOOLQESQOQMLGRLIPEKQILSDQLKQVQNSLHRSLSLTALKALEKELARQO 121
 QY 581 LREQLDEVERETRSKIQEIDVFNNOIKELREIHSKQQLQKQSLERARLKQEKQEKSLIE 640
 DB 122 LRDQLDEVERETRSKIQEIDVFNNOIKELREIHSKQQLQKQSLERARLKQEKQEKSLIE 181
 QY 641 LEKQEDARVROERKQKMLEHVQOE-EGPRPKPHEERLKLREDSVRKKEAEERKPEM 699
 DB 182 LEKQEDARVROERKQKMLEHVQOEDEHQRPKLHEERLKLREDSVRKKEAEERKPEM 241
 QY 700 QDKOSHLFPHQEPAPKATQAPWSTTEKPLTISAQESVAVVYRYALYFESSHSEIT 759
 DB 242 QDKOSHLFPHQEPAPKATQAPWSTTEKPLTISAQESVAVVYRYALYFESSHSEIT 301
 QY 760 QGEDIWVDESQTEGEGMLGELIKGTGMPNANYAEKIPENEVPTPAKVTDTITSAAPK 819
 DB 302 QGEDIWVDESQTEGEGMLGELIKGTGMPNANYAEKIPENEVPTPAKVTDTITSAAPK 361
 QY 820 LALRETPAPLPTSSPSTTPNNWADPSSSTWSSNKEKETNNWDMWAQPSITVPSAQO 879
 DB 362 LALRETPAPLPTSSPSTTPNNWADPSSSTWSSNKEKETNNWDMWAQPSITVPSAQO 421
 QY 880 LKORSAPFTATAGSSPSPVLGGEGKEVGLQALTPMRAKKNHNFNKSVDITVLEQO 939
 DB 422 LKORSAPFTATAGSSPSPVLGGEGKEVGLQALTPMRAKKNHNFNKSVDITVLEQO 481
 QY 940 DMWPFEEVQOQKGMFKFSYVKLISGPRKSTSIDTGPTEPSASLKVASPAKPAIIGEE 999
 DB 482 DMWPFEEVQOQKGMFKFSYVKLISGPRKSTSIDTGPTEPSASLKVASPAKPAIIGEE 491

QY 1000 FIAMTYESSEQGLTFPOGDVIVVTKKQDGMWMTGVBKSGVPPSNVYRLKDSSESGTA 1059
 DB 492 FIAMTYESSEQGLTFPOGDVIVVTKKQDGMWMTGVBKSGVPPSNVYRLKDSSESGTA 551
 QY 1060 GKTGSLGKKPEIAQVIAATGATPQTLTAPQQLILIRKQNGGWEGLQARQKQIQ 1119
 DB 552 GKTGSLGKKPEIAQVIAATGATPQTLTAPQQLILIRKQNGGWEGLQARQKQIQ 593
 QY 1120 WFPANYKLLSPGTSKITPTELPTAVQAVQVIGMDTYQNDDELAFSGQIINVTN 1179
 DB 594 -----VCQVIGMDTYQNDDELAFNGQIINVTN 623
 QY 1180 KEDPMMKGEVAGVGLPSPNVKLTITMDPSQO 1213
 DB 624 KEDPMMKGEVAGVGLPSPNVKLTITMDPSQO 657
 RESULT 9
 ID AAO17881 standard; Protein; 1681 AA.
 XX AAO17881;
 AC 22-AUG-2002 (first entry)
 DT
 XX Allergic disease examination method related human protein.
 DE Allergic disease; allergy; antiallergic; intersecin 2; eosinophil;
 KW atopic dermatitis; human.
 KM
 XX Homo sapiens.
 OS
 XX WO200233122-A1.
 PN
 XX 25-APR-2002.
 PD
 XX 11-OCT-2001; 2001WO-JP08937.
 PF
 XX 13-OCT-2000; 2000JP-0314093.
 PR
 XX (GENO-) GENOX RES INC.
 PA (NIGE-) JAPAN GEN NAT CHILDREN'S HOSPITAL.
 PA (EISA) EISAI CO LTD.
 XX Sugita Y, Hashida R, Ogawa K, Obayashi M, Nagaen T, Saito H;
 PI Takahashi E;
 DR MPI; 2002-372313/40.
 DR N-PSDB; AAL47247.
 XX
 XX Method for examining allergic diseases by differential display of
 PT intersecin 2 gene showing different expression particularly
 PT significant increase in eosinophils in patients
 XX
 PS Disclosure; Page 78-86; 90pp; Japanese.
 PS
 CC The present invention relates to a method for examining allergic diseases
 CC with intersecin 2 gene or a gene with equivalent function of intersecin
 CC 2 as an indicator gene, which comprises determining the expression level
 CC of the gene in the eosinophils in a patient, and comparing the expression
 CC level with that in the eosinophils of a healthy individual. The method is
 CC for examining allergic diseases, particularly atopic dermatitis, which is
 CC also applicable in screening candidate compounds for remedies. The
 CC present sequence is a human protein described in the exemplification of
 CC the invention.
 CC
 SQ Sequence 1681 AA;
 Query Match 49.3%; Score 3091.5; DB 23; Length 1681;
 Best Local Similarity 52.1%; Pred. No. 1.2e-189;
 Matches 656; Conservative 187; Mismatches 274; Indels 143; Gaps 29;
 QY 1 MAQPTPGGSLDVAIVTEBRKHDQFLSLKPTAGTITGQANFFQSGLPQPVLAQ 60

```

||||| 13 MAQFTANNGGNMMAITSEERTKHDQFDNKLPSGGYITGQANFLQSLGPAFVLA 72
||||| 61 IVALADNNNGDMDOVESIAMKLIKLYQYQLPSTLPVWKQOPV--AISAPAFIG 118
||||| 73 IVALSDLNKDKGMDQOEFSIAMKLIKLYQYQLPSTLPVWKQOPV--AISAPAFIG 131
||||| 119 GIASMPPLTAVAPVPMG-----SIPVVGMPPLVSSVPPAAPPVLANGAPV 165
||||| 132 ---SNPNUSIPQPLPPAAPITSLSSATSGTNLPLMLPPLVPSVSTSLP---NGTASL 185
||||| 166 IQPLPAFAHPAATWPKSSSFS--RSGPGSQLNTKLQAKQ--FDVASAP----- 211
||||| 186 IQPLP-IPYSSSTLPHGSSYSILMGWFG--GASIQXQSLDLDCSSSTSTASLGSNS 241
||||| 212 ----AAEAVAVPOSSSLKTRQLFNSHDKTMSGHLTGPQARTIMQSSLSQALASITWISD 267
||||| 242 PKGTISEAVAVPOPTSLKTRQKENTLTKSMGSLSGFOARNALLQSNLSQTLATITWISD 301
||||| 268 IDQDKLTAEFFILAMHLIDVAMSGOPLPPVLPPEYIPSPFRVAVSGSMVSSSYDQ 327
||||| 302 VDGQQLKAEFFILAMHLIDVAMSGOPLPPVLPPEYIPSPFRVAVSGSMVSSSYDQ 353
||||| 328 RLPEPSSSEDOQPEKULVTPEDKKRNFERSGVELEKRRQALQEQKQEQERLAQLER 387
||||| 354 TLPSYQKQV-EEEPQKULVTFEDKKRNFERSGVELEKRRQALQEQKQEQERLAQLER 412
||||| 388 AEQEKERERQEQAKQLEKQLEKQLEKQLEKQLEKQLEKQLEKQLEKQLEKQLEKQLE 447
||||| 413 EEMERKORLEQEQWKQLEKQLEKQLEKQLEKQLEKQLEKQLEKQLEKQLEKQLEKQLE 472
||||| 448 WERNRROELNORNEQEGTVVLKARKTLEFELKQLEKQLEKQLEKQLEKQLEKQLEKQLE 507
||||| 473 WERTROELNORNEQEGTVVLKARKTLEFELKQLEKQLEKQLEKQLEKQLEKQLEKQLE 532
||||| 508 EIESTNKSRELIAETIHLQOQLOESQOQMLGRLPEKQILSDQLQOVQ--QNSLHSDSLT 566
||||| 533 ELEVDKQODLEIMEIKQLOQLOEQYQNKLIYLVPEKQILNRIKQWQNSNTPDSGVSL 592
||||| 567 LKRALBAKELAQLEQLEDEVERTRSKQLEIDVFNNOLEKRLREHSHQQLQOKRSLBA 626
||||| 593 HKKSLEKEBELC-QRKEQLDALEKETASKLSEMDFNQOLKELRETVNQQALQALQYKI 651
||||| 627 ARLKQEQERKSLSELEKQK---EDQORVORERDQOMLEHVOEQEOPRRKHEBRLK 682
||||| 652 KQDKLEIKERKLELMQKKLEDEAKRKAQKQKEMIKENLKEE-----EKQKLOE 705
||||| 683 EDSVRKKEAEERAKPEMDQKSLFPHQEPAKLATQAWSTTEKPLTISAQESVK--- 739
||||| 706 EKTQEKIQEERKAEKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQK 740
||||| 740 --VVTYRALYPPESRSHDEITIQPGDIWAVDSQGEPEWLGELKGTGMPAYAEKI 797
||||| 741 SVLVNTRALYPPEARNHDEMSFNSGDIQVDEKTYGEPEWLGSGQNGMPPCYVEEM 800
||||| 798 P--ENEVPPAPVNDUTSAPAPKALRETAPLVPTSESPSTTPNNMADFSTWSSSN 855
||||| 801 PSENE-----KAVSPKRAL-----LPTVLSLATS-----TSSEPLSN 835
||||| 856 EKPEIDMDMTWAAQPSLTVPSAGOLRQSAFTPATATGSSPSVPLQGEKVEGLQALY 915
||||| 836 QASAVTDYQV-VSFSNLTVNTSMQ--KSAFRTYSPG-SVPHHQGGQVENVLKAQALC 891
||||| 916 PPRAKDNHLNFKSDVITVLEQODMMWGEVQOGKMPKSYVKLISPPVKSTSIDTG 975
||||| 892 SMTAKDNHLNFKSDVITVLEQODMMWGEVQOGKMPKSYVKLIPSEVK----- 944
||||| 976 PPSAPSLKRVAS--PAKPAIPGEFFIMTYEESQGLTFQOGSDVIVTKQDQWMT 1033
||||| 945 -REEPALYAAVVKKTSAAVGESEYIALYPPSSVEPDLTFTBEELIVTQKQGEWMT 1003
||||| 1034 GTVGDKSGVPSNYYVLKDSSESGTAGTSGLQKKEIQAVALASVATGPEQTLAPQL 1093
|||||

```

```

Db 1004 GSIGDRSGIFPSPNVYKPDQESFGSASKSGASNKKPEIAQVTSAYVAGSEQLSLABQOL 1063
Qy 1094 ILIRKKNPGMWESELOARKKROIGFPPANVYKLSPGSKITPTLPTAVGPANCOV 1153
Db 1064 ILIRKKNPGMWESELOARKKROIGFPPANVYKLSPGSKITPTLPTAVGPANCOV 1116
Qy 1154 IGMVDTYACNDDELAFSGQIINVLNKEDEPDMMWKGESGQVGLFPSPNVYKLTMDPSQ 1213
Db 1117 IGMVDTYACNDDELAFSGQIINVLNKEDEPDMMWKGESGQVGLFPSPNVYKLTMDPSQ 1176

RESULT 10
AAV32157
ID AAV32157 standard; Protein; 648 AA.
XX
AC AAV32157;
XX
DT 01-FEB-2000 (first entry)
XX
DE Human SH3D1A protein.
XX
KW SH3D1A, human; Down's syndrome; leukaemia; cancer;
KW megakaryocytic abnormality; myeloproliferative disorder;
KW platelet disorder; neural disorder; thrombocytopenia;
KW haematopoietic disorder; cognitive dysfunction; microcephaly;
KW lissencephaly; colpocephaly; holoprosencephaly; diagnosis; therapy.
XX
OS Homo sapiens.
XX
PN M09953062-A2.
XX
PD 21-OCT-1999.
XX
PF 16-APR-1999; 99WO-US08371.
XX
PR 16-APR-1998; 98US-0082007.
XX
PA (CEDA-) CEDARS SINAI HEALTH SYSTEM.
XX
PI Korenberg JR, Chen X;
XX
DR WP1; 1999-633829/54.
XX
DR N-PSDB; AAZ34573.
XX
PT Nucleic acid from the human SH3D1A gene and its products, useful for
PT leukaemia -
PS
PS Claim 14; Fig 13; 99p; English.
XX
XX This sequence represents the protein encoded by the human SH3D1A
XX CDNA clone 5 (see AAZ34573). SH3D1A contributes to the development
XX of platelets and the pathogenesis of leukaemias, both in general
XX and in particular those involving the megakaryocytic lineage. The
XX gene maps to the small candidate region for low platelets on
XX chromosome 21. The protein includes SH3 domains and EH domains,
XX both associated with protein-protein interactions and the latter
XX with maintenance of the cytoskeleton. At least 3 isoforms of
XX SH3D1A exist (see AAV32154-58). The invention provides methods for
XX the diagnosis and treatment of megakaryocytic abnormality,
XX myeloproliferative disorder, platelet disorder, acute leukaemia,
XX neural disorders, thrombocytopenia, platelet disorder on
XX chromosome 21, low platelets in deletion for 21, association of
XX gains in chromosome 21 with leukaemias, neural abnormalities,
XX dysfunctions and disorders including brain malformations and
XX corresponding cognitive dysfunctions, microcephaly, lissencephaly,
XX cell and colpocephaly. Methods are also provided: for suppressing
XX cells unable to regulate themselves; screening for a somatic
XX alteration in the SH3D1A gene; monitoring the progress and
XX adequacy of a treatment; monitoring tumour risk progress or
XX megakaryocytic abnormality; myeloproliferative disorder,
XX haematopoietic disorder, platelet disorder or leukaemia; and
XX screening of drugs for cancer therapy.

```

XX Sequence 648 AA;
SQ Query Match 49.0%; Score 3070.5; DB 20; Length 648;
Best Local Similarity 94.1%; Pred. No. 7,4e-189;
Matches 609; Conservative 20; Mismatches 17; Indels 1; Gaps 1;
QY 1 MAQFPFGSLDVAITVEERAKHDQFLSLKPIAGFTTGDQARFFPFGSGLPQVLAQ 60
DB 1 MAQFPFGSLDVAITVEERAKHDQFLSLKPIAGFTTGDQARFFPFGSGLPQVLAQ 60
QY 61 IVALADNMNDGRMDQVEFSIAMKLIKIKLQGYQLPSTLPPVMKQPVASISAPAFGIGI 120
DB 61 IVALADNMNDGRMDQVEFSIAMKLIKIKLQGYQLPSTLPPVMKQPVASISAPAFGIGI 120
QY 121 ASMPPLTAAPVPMGSIPIVVGMSPLVSSVPPAAPPPLANGAPPVQLPAPAHAPATMP 180
DB 121 ASMPPLTAAPVPMGSIPIVVGMSPLVSSVPPAAPPPLANGAPPVQLPAPAHAPATMP 180
QY 181 KSSFSRSGSGSQLNTKLQKAGSFDVASAPPAEMVAVPOSSRLKTYROLFNSHDKTMSGH 240
DB 181 KSSFSRSGSGSQLNTKLQKAGSFDVASAPPAEMVAVPOSSRLKTYROLFNSHDKTMSGH 240
QY 241 TGPQARTILMOSSLPOAQLASINWLSLDDQDKLTAEFLIAMLIDVAMSGPLPPLVP 300
DB 241 TGPQARTILMOSSLPOAQLASINWLSLDDQDKLTAEFLIAMLIDVAMSGPLPPLVP 300
QY 301 PEYIPSPFRVNSGSGMSVYSSSVQRLPEEFSSEDEQOP-EKULPVTFEDKKRNFER 359
DB 301 PEYIPSPFRVNSGSGMSVYSSSVQRLPEEFSSEDEQOP-EKULPVTFEDKKRNFER 359
QY 360 GSVELEKRRQALLQQRKEGERLAOLERAEOEKERERBOEAKROLELEKOKERE 419
DB 360 GSVELEKRRQALLQQRKEGERLAOLERAEOEKERERBOEAKROLELEKOKERE 419
QY 420 ROREERREKREIERREAKLELROLEWRNRROELINQKNEOEGTVLTKARKTLEF 479
DB 420 ROREERREKREIERREAKLELROLEWRNRROELINQKNEOEGTVLTKARKTLEF 479
QY 480 ELBALNDKKGHLEGGKQDIRCRLATROEIESTNKREIRIAITHTLQOOLQESQMLGR 539
DB 480 ELBALNDKKGHLEGGKQDIRCRLATROEIESTNKREIRIAITHTLQOOLQESQMLGR 539
QY 540 LIPKQILSDOKQVOONSILHRPSLTLTKRALEAKLAAQOLDEQIDVERETRSLOEI 599
DB 540 LIPKQILSDOKQVOONSILHRPSLTLTKRALEAKLAAQOLDEQIDVERETRSLOEI 599
QY 600 DVFNQOLKELEIRIHSKQOLQKRSLEAARLKQKOEKRSLELEKOKE 646
DB 600 DVFNQOLKELEIRIHSKQOLQKRSLEAARLKQKOEKRSLELEKOKE 646
QY 601 DIFNQOLKELEIRIHSKQOLQKRSLEAARLKQKOEKRSLELEKOKE 647
DB 601 DIFNQOLKELEIRIHSKQOLQKRSLEAARLKQKOEKRSLELEKOKE 647
RESULT 11
AAM79199
ID AAM79199 standard; Protein; 1697 AA.
XX AAM79199;
AC AAM79199;
DT 06-NOV-2001 (first entry)
DE Human protein SEQ ID NO 1861.
XX
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation.
OS Homo sapiens.
XX
XX MO200157190-A2.
PN 09-AUG-2001.
XX
XX

PF 05-FEB-2001; 2001WO-US04098.
XX
XX 03-FEB-2000; 2000US-0496914.
PR 27-APR-2000; 2000US-0568875.
PR 20-JUN-2000; 2000US-0598075.
PR 19-JUL-2000; 2000US-0620325.
PR 01-SEP-2000; 2000US-0654936.
PR 15-SEP-2000; 2000US-0663561.
PR 20-OCT-2000; 2000US-0693325.
PR 30-NOV-2000; 2000US-0728422.
XX
XX (HYSE-) HYSSEQ INC.
XX
XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao QA, Wang D, Wang J, Zhang Y, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
XX WPI; 2001-476283/51.
DR N-PSDB; AAK52332.
XX
PT Nucleic acids encoding polypeptides with cytokine-like activities,
XX
XX useful in diagnosis and gene therapy -
XX
XX Claim 20; Page 4246-4249; 6221pp; English.
XX
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAK80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activity/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX
XX SQ Sequence 1697 AA;
QY Query Match 48.9%; Score 3068; DB 22; Length 1697;
Best Local Similarity 51.0%; Pred. No. 4.1e-188;
Matches 656; Conservative 187; Mismatches 274; Indels 170; Gaps 30;
QY 1 MAQFPFGSLDVAITVEERAKHDQFLSLKPIAGFTTGDQARFFPFGSGLPQVLAQ 60
DB 2 MAQFPFGSLDVAITVEERAKHDQFLSLKPIAGFTTGDQARFFPFGSGLPQVLAQ 61
QY 61 IVALADNMNDGRMDQVEFSIAMKLIKIKLQGYQLPSTLPPVMKQPV--AISSAPAFGIG 118
DB 62 IVALADNMNDGRMDQVEFSIAMKLIKIKLQGYQLPSTLPPVMKQPV--AISSAPAFGIG 120
QY 119 GIASMPPLTAAPVPMG-----SIPVVGMSPLVSSVPPAAPPPLANGAPPV 165
DB 121 ---SMPNLSIPPLPAAPITSLSSATSGTNLPLMPPLPVPVSTSLP---NGTASL 174
QY 166 IQPLPAFAHAPATWPKSSSFS--RSGPGSQLTKLQKAGS-FDVASAP----- 211
DB 175 IQPLP-IPYSSSTLPHGSSYSILMWGFG---GASIQKQSLDLSSSSSTSTASLGS 230
QY 212 ----AAEMAVPOSSSLKTRQLFNSHDKTMSGHITGPQARTILMOSSLPOAQLASINWLS 267
DB 231 PKTGTSEMAVPOPTRLKTRQKENTLDKMSGYLSFGQARNALQNTLSQTQATATWTLAD 290
QY 268 IQDQKLTAEFLIAMLIDVAMSGOPLPVVLPPEYIPSPFRVNSGSGMSVYSSSVQ 327
DB 291 VPDGQQLAEFLIAMLIDVAMSGOPLPVVLPPEYIPSPFRVNSGSGMSVYSSSVQ 342
QY 328 RLPEEFSSEDEQOPKULPVTFEDKKRNFERSVLEKRRQALLQQRKEGERLAOLER 387
DB 343 TLPSYQXQG-EPEQPKULPVTFEDKKRNFERSVLEKRRQALLQQRKEGERLAOLER 401

```

388 AEOEKEREROEAKROLLEKOLEKORLEEROREERREKREERREAKRELEOROLE 447
402 EEMERKOLELOEKOLEKOLEKORLEEROREERREKREERREAKRELEOROLE 461
448 WERNROELNORNEOGTVLAKARKTLEFELEBALNDKHOLEGKLODIRCLATORO 507
462 MERIRGOLLNOKNEGEIIVRLNSKKNLHLEBALNGKQISGRLOVRLKQKQKT 521
508 EIESNKRERELRIATITLHQOOLQESQOMGLRIEKOILSDOLQOV-ONSILHDSILT 566
522 ELEVLDKOCDEIMEIKOLOELQEQNKLYLVEPEKOLMERIKMOPNSVPGVSLT 581
567 LKBALEAELARQOLREOLDEVERETREKLOEIDVFNNO----- 605
582 HKKSLEKEELC-QRLKEQDLDALEKETASKLESDMSFNQOLKGMNDVLOCLLSL 640
606 -----LKELEREINSHKQOLKORSLAARLKQEKORERKLEKOK---EDAORVQER 655
641 NNLFLLLEKELRETVNTQOLALEQYKIKRDKLKEIERKRLLEMKKLEDEAKRQK 700
656 DKOMLEHVQOEOPPRKPRHEEDRLKREDSVKKAERAKPEWODKQSLFHPHQEP 715
701 ENLMKERNRKEBE-----EKOKRLOEKETOEKLOEBERKKEKOKRDXD----- 744
716 LATQAPMSTTEKGLTISAOSVK-----VYYRALYPFESRSHDEITOPGDIVM 770
745 -----TLKAEKREKRETSASVLYNRYALYFEAENHDEMSFNSDIILQVEK 789
771 QTEBERGMLGELKGTGWFNPANYAEKIP--ENEVPTPAKPYDILTASAPKILARET 828
790 TVGEKGMVLYGSGFGKGFPCVYVEKMSENE-----KAVSPKKAL----- 831
829 LPVTSSSESTTPNMADPSSSTWPSSENEKPEETDMMDTAAOBSLTVPSAGOLRQ 888
832 LPPTVSLATS-----TSSEPLSSNQPAVTDYON--VFSNLTVTNTSMQ--KKS 881
889 ATATSSSPVLYGGEKEVGLQAOALYPMRAKONHLNPNKSDVTTVLEQODMMPGE 948
882 TVSPG-SVSPHGOQOVENLKAQMLCSWTAKKDHNLFSKHDITVLEQENMMPGE 940
949 GQKGMFPKSYVYLISGPVRSKTSIDTGPESPAKLRVAS--PAKPAIPGEEPTAMY 1006
941 GGRGMFPKSYVYLISGPVRSKTSIDTGPESPAKLRVAS--PAKPAIPGEEPTAMY 992
1007 ESSSEGDLTFQOGDVIYVTKKODMWTGTGDKSGVPSPSYRRLKDSGSGTAGT 1066
993 SSVEPGDLTFEGEELIVTOKGEMWMTGISIGRSGIFPSNYVKPPOQESFGS 1052
1067 KKPEIAOVIASVAAATGPEQLTLAPGOLLIRKONGGMEGLQARGGKROIGMFP 1126
1053 KKPEIAOVTSAVAVSGESQLSLAPGOLLIRKONSGMGGLOARGGKROIGMFP 1112
1127 KLSSEGTSKITPTLEPKTAVOPAVQVIGMDYTAQNDELAFSGQIINVLNKED 1186
1113 KLGSSSRATPAHP-----VCQVIAWYDAAANNDELSEFSKQILINWKNOD 1165
1187 KGEVSGQVGLFPSPNYVKLTMDPSQO 1213
1166 OGEINGVTGLFPSPNYVKMTTDSPOQ 1192

```

```

KW regulation; actin cytoskeleton; detection; cancer; infection;
KW EH-domain and SH3-domain regulator of endocytosis; anticancer;
KW antiproliferative; antiviral.
XX
XX Mus sp.
XX WO955728-A2.
XX
XX 04-NOV-1999.
XX
XX 27-APR-1999; 99WO-CA00375.
XX
XX 27-APR-1998; 98CA-2230201.
XX 05-FEB-1999; 99US-0118739.
XX (HSCR-) HSC RES & DEV LP.
XX
XX Egan SE, Wang W, Sengar A;
XX MPI; 2000-052802/04.
XX N-PSDB; AA239010, AA239011.
XX
XX New nucleic acid encoding Ees1 and 2 proteins, involved in regulation
XX of endocytosis, used e.g. for treating cancer or preventing viral
XX infection.
XX
XX Claim 33; Page 48; 99pp; English.
XX
XX The present sequence represents mouse Ees2. The present invention
XX specifically describes mammalian Ees1 and 2 proteins (I) and their splice
XX variants (Ese = EH-domain and SH3-domain regulator of endocytosis). (I)
XX are involved in regulation of clathrin-mediated endocytosis (as a complex
XX with Ees1s protein), vesicular trafficking and actin cytoskeleton.
XX Generally (I) (or its (ant)agonists, mimetics, fragments and inactive
XX mutants); (II)-specific antibodies (Ab); sequences antisense to the (I)
XX polynucleotide; agents that downregulate expression of Ees genes or
XX antagonists of an Ees binding partner are used to treat diseases
XX associated with undesirable endocytosis and resulting changes in cellular
XX function. Particularly overexpression of Ees1 is used to block
XX clathrin-mediated endocytosis in vivo or in cell cultures, while
XX administration of (I) is used to promote endocytosis of selected cells.
XX (Ant)agonists of (I) or Ab are used to suppress abnormal proliferation of
XX cells that can be stimulated to proliferate by a growth factor receptor;
XX and similar compounds (also inactive Ees mutants) can be used to prevent
XX viral infection. Endocytosis may also be regulated, in vivo or in cell
XX cultures, by forming an Ees-Ees1s complex, then binding dynam to the
XX complex. Generally conditions that can be treated include cancer;
XX abnormal cell division or migration; viral infection; or abnormal
XX receptor signalling, tissue development or synaptic transmission.
XX
XX Sequence 1197 AA;
SQ
Query Match 48.6%; Score 3049.5; DB 21; Length 1197;
Best Local Similarity 52.4%; Pred. No. 3.9e-187;
Matches 658; Conservative 177; Mismatches 275; Indels 145; Gaps 32;

```

Db 230 SPKGTSEWAVPQPSRLKTRKQKNSLDKMGVSLSGFQANALQSNLSQTLATITWTLA 289
 QY 267 DIDDOCKLTAEEIILMHLIDVAMSGOPLPVLPVPIPPSPFRVAVSSGSMVSISSSD 326
 Db 290 DIDGDDQLAAEEITLMMHLIDVMAKQPLPLTLPPLVPSFR-----GGKQV---DSVN 341
 QY 327 ORLPPESSSEDOQPEKLPVTFEDKKRENFERSVLEKRRQALHEOORKEOERLAQJE 386
 Db 342 GILPSQKQ-EEEPQKLPPTTEDGRKANVERGMELEKRRQVLMEOQREERAKQKE 400
 QY 387 RAEOEKREBERQOAKROLELEKOLEKOLEKOLEKOLEKOLEKOLEKOLEKOLEKOLE 446
 Db 401 KEEMERKORELOQEWKQOLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLE 460
 QY 447 EWEHRNROELNORNEQESTVVLKARKTLEFELALNDKQKQLEKQLEKQLEKQLEKQ 506
 Db 461 EWEHLRROELLSQKTEOEDIVRLSSRKSLHLEBAVNGKQOISGRUDVQIRKQOK 520
 QY 507 OEIESTNKSRELRIRAEITHLQOOLQOESQOMLGRILPEKQILSDQLKOVQONSJHR 566
 Db 521 TELFVLDKQCDLEIMTKOULOQELKEYQNKLIYLBKQULNERIKQMOLSTPDGIGSI 580
 QY 567 LKRALAEKELARQOLREQLEDEVERETRSKLOEIDVENNOULKELEIHSKQOLQOK 626
 Db 581 LHKSSSEKEELCORLKEQDLEKETAASKLSEMSFNNOULKELEIHSKQOLQOK 640
 QY 627 ARLKQOEORKSLE-LEKQ---EDAKRQOEKQOMLEHVQOEOPPRPKPHEEDRLK 682
 Db 641 KQDKLEIERKRLKQOQKLEDEARAKQKQKEMLRSEIRKEE-----EKQRLQJE 694
 QY 683 EDSVRKEAEERAKPEMODKQSLFPHQEPALATQAPMSTTEKPLTISAQESKVYV 742
 Db 695 EKSQDKTQSEER---KAKAQSE-----TASA-----LVN 721
 QY 743 YRALYFESRSHDEITIQPDIVMDESQTEBGMVLGELKGTGMFPANYAEKIPENY 802
 Db 722 YRALYFPEARNDHEMFSSGDIQVDEKTVGEKGMVLYGSHQKGFQPCVYEVKLSSE 780
 QY 803 PTPAKVNTDLSAPAKALRETPALPVTSSEPTSTPNMADPSSSTWPSSENEKETDN 862
 Db 781 -----KALSFRKALLPPTVSLSATSTIS-SSPPASVTDYHNV--SFSNLTVNT-- 824
 QY 863 WPTMAQPSLTVPSAQOLRQSAFTPATATGSSPSVPLQGGVEGLQALQALYPMRAXD 922
 Db 825 --TW-----OQKSAFTTVPFG-SVSPFHGGQQAVENTLKQALCSWPAKX 867
 QY 923 NHIENFKSDVITVLEQOQDMWFMGEVQOKGMFPKSYVKLISG-PVRS--TSIDGPTES 979
 Db 868 NHIENFKSHVITVLEQOQDMWFMGEVQOKGMFPKSYVKLIPGNEVQORGEPAALYAVTK 927
 QY 980 PASLKVASPAKPAIP-GEETIAMTYESSSEGQDLTPOQDVIIVTKKDGDMWTGVD 1038
 Db 928 PHS---TAAVPVISTAPVGEDYIALYSYSVEBPDGLTFEGEILVLTQKQGEWMTSIS 984
 QY 1039 KSGVPSNVYRLKDSGSGTAGTGLKKEPTEIAOVITASYAAGPRLTLAPQQLILIK 1098
 Db 985 RTGIFPSNVYRPPQDQENFGNASKSGASNNKPBTAQVTSAAAGSTQSLSLAPQLILIK 1044
 QY 1099 KNPFGWEGELQARKKROIGMFPANYVKLISPTSKITPTTELPTKTAPOVAVQVIGMD 1158
 Db 1045 KNTSGWQGELOQARKKROIGMFPANYVKLISPTSKITPTTELPTKTAPOVAVQVIGMD 1097
 QY 1159 YTAQNDDELAFSGQIITVANKEDPMMKGEVSGVQGLPSPNYVKLITMDPSQ 1213
 Db 1098 YMANNEDELIFSQGLINVMKDDPMMQGETINGLTGLFSPSNYVKMTTSDPSQ 1152

XX 28-FEB-2000 (first entry)
 DT Mouse Ees2l protein sequence.
 DE Mouse; murine; Ees1; Ees2; endocytosis; vesicular trafficking;
 KW regulation; actin cytoskeleton; detection; cancer; infection;
 KM EH-domain and SH3-domain regulator of endocytosis; anticancer;
 KM antiproliferative; antiviral.
 OS Mus sp.
 PN MO9955728-A2.
 PD 04-NOV-1999.
 PF 27-APR-1999; 99MO-CA00375.
 PR 27-APR-1998; 98CA-223021.
 PR 05-FEB-1999; 99US-0118739.
 PA (HSCR-) HSC RES & DEV LP.
 PI Egan SE, Wang W, Sengar A;
 DR MPI: 2000-052802/04.
 DR N-PSDB; AA239026, AA239027.
 XX New nucleic acid encoding Ees1 and 2 proteins, involved in regulation
 PT of endocytosis, used e.g. for treating cancer or preventing viral
 PT infection -
 PS Claim 33; Page 69-70; 99p; English.
 CC The present invention specifically describes mammalian Ees1 and 2
 CC proteins (I) and their splice variants (Ese = EH-domain and SH3-domain
 CC regulator of endocytosis). (II) are involved in regulation of clathrin-
 CC mediated endocytosis (as a complex with Eps15 protein), vesicular
 CC trafficking and actin cytoskeleton. Generally (I) (or its (ant)agonists,
 CC mimetics, fragments and inactive mutants); (I)-specific antibodies (Ab);
 CC sequences antisense to the (I) polynucleotide; agents that downregulate
 CC expression of Ees genes or antagonists of an Ees binding partner are
 CC used to treat diseases associated with undesirable endocytosis and
 CC resulting changes in cellular function. Particularly overexpression of
 CC Ees1 is used to block clathrin-mediated endocytosis in vivo or in cell
 CC cultures, while administration of (I) or Ab are used to suppress abnormal
 CC proliferation of cells that can be stimulated to proliferate by a growth
 CC factor receptor; and similar compounds (also inactive Ees mutants) can be
 CC used to prevent viral infection. Endocytosis may also be regulated, in
 CC vivo or in cell cultures. By forming an Ese-Eps15 complex, then binding
 CC dynamitin to the complex. Generally conditions that can be treated include
 CC cancer; abnormal cell division or migration; viral infection; or abnormal
 CC receptor signalling, tissue development or synaptic transmission. The
 CC present sequence represents mouse Ees2l protein sequence.
 SQ Sequence 1658 AA;
 Query Match 48.6%; Score 3049.5; DB 21; Length 1658;
 Best local Similarity 52.4%; Pred. No. 6.1e-187;
 Matches 658; Conservative 177; Mismatches 275; Indels 145; Gaps 32;
 QY 1 MAQFPFPGSGSLDVAITVEERAKIDQFLSKPIAGFTGQARNFFQSGLPQVLAQ 60
 Db 1 MAQFPFANNGGNNMAITSEERTKDKQFDNKPSCGYITGQARTFFLOGLPAPVLA 60
 QY 61 IVALADNMNDGMDQVPSIAKLIKIKLQGVQLPSTLPVVKQOPV--AISAAPFGIG 118
 Db 61 IVALSLDKDGKMDQVPSIAKLIKIKLQGVQLPSTLPVVKQOPV--AISAAPFGIG 119
 QY 119 GIASMP-----PLTAVAPV--PMG-----SIPVVGSPPLVSSVPPAAPPPLANGAP 164
 Db 120 ---SMPNLSIHQPLPVPVAPVATPPLSSATSGTSTIPPLMMPAPLVPSTSLP---NGTAS 173

```

165 VIOPLPAFAHAAATWPKSSSFS--RSGPSQNLTKLOKAO--FDVASAP----- 211
174 LIQPL-STIPSSSTIPHASSTYSLMMGFG--GASIQAOQILIDGSSSTSSSTASLSCN 229
212 -----AAEMAVPOSSRLKYROLFNSHDKTMSGHLTGPARITLMQSSLPQOALASIVNLS 266
230 SPKGTISEMAVQPQSRKLKYRQKFNSLDKMGSGYLSGFGAPANALQSNLSQQLATITWLA 289
267 DIDQDKLTAEFFILAMHLIDVANGQPLPPLPPEYTPPSFRRRSSGMSYSSSSVD 326
290 DIDGGQKABEFILAMHLIDMAKAGQPLPLPPELVPSFR-----GGQV--DSVN 341
327 QRLPEPSEBQOEKLPVTFEDKKENFERGVELEKROALLQOEORKEORLAE 386
342 GTLPSYQTO--EEFQKLPVTFEDKKENFERGVELEKROALLQOEORKEORLAE 400
387 RAEORKEORKEORKEORKEORKEORKEORKEORKEORKEORKEORKEORKEOR 446
401 KEEMERKQRELQOEQWKQLEKQLEKQLEKQLEKQLEKQLEKQLEKQLEKQLEK 460
447 EEMERKQRELQOEQWKQLEKQLEKQLEKQLEKQLEKQLEKQLEKQLEKQLEK 506
461 EEMERKQRELQOEQWKQLEKQLEKQLEKQLEKQLEKQLEKQLEKQLEKQLEK 520
507 OEIESTNSRELRLAETHTLOOQLOESQOMGLPPEKQIILSDOLKQVQNSLHRSILT 566
521 TELEVLQOCDEIETIKOLOEELKEYONKLYLPEKQLEKQLEKQLEKQLEKQLEK 580
567 LKRALAEKELARQOLREQLDEVERETRSKLOEIDVFNQOLKELEHSHKQOLQORSLEA 626
581 LHKSSSEKELCQRLKEQDLALEKETASKLSEMDSEFNQOLKELEHSHKQOLQORSLEA 640
627 ARLKQOEORKEORKEORKEORKEORKEORKEORKEORKEORKEORKEORKEOR 682
641 KRDQKLEKQLEKQLEKQLEKQLEKQLEKQLEKQLEKQLEKQLEKQLEKQLEK 694
663 EDSVKKAEERAKEMODKOSRLPHPHQEPAKLATQAPWSTEGPLTISAGQSVKVVY 742
695 EKSQDKTOEER--KAERKQSE-----TASH-----LVN 721
743 YRALYPPESRSHDEITIQGDIWVDESQTEBPGWLAGELKGTGMPFAPYAKIENEV 802
722 YRALYPPESRSHDEITIQGDIWVDESQTEBPGWLAGELKGTGMPFAPYAKIENEV 780
803 PTPAKPVTDLTISAPKALARETAPLPVTSSEBPTNNMADBSSTPSSSEKPEPTDN 862
781 -----KALSPKRALPPTVLSLSTSTSS--SQPPASVTDYHNV--SFSNLTVNT-- 824
863 WDTMAQPSLTVPSAGLQORSAPFPATATGSPSPVVGQEKVAGLQOALYPMRAKQD 922
825 --TW-----QKSAFRTTVSPG--SVSPFHQGOAVENLQOALYPMRAKQD 867
923 NNLNFKNSDVTITVLEQODMMWFGVQOGKGFEPKSYVLISG--PYRKS--TSIDTGPES 979
968 NNLNFKNSDVTITVLEQODMMWFGVQOGKGFEPKSYVLISG--PYRKS--TSIDTGPES 927
980 PASLKRVASPAKPAIP--GEFIAMTYTESSEQGLDTQOGQDVIVYTKKQDMWTGVGD 1038
928 PTS--TAYPTVSTAYPVGEDTIALYSSVPEGLTFEGEELIVTQKDESMWTGSIGE 984
1039 KSGVPSNYVVLKQSESGTAKGSLGKKEPILQIVTASVATGSEOLTLAAGQILIRK 1098
985 RTGLFPSNYVVRPKQDENFGANSKSGANSKKEPILQIVTASVATGSEOLTLAAGQILIRK 1044
1099 KNPQGMWEGELQARQKROIGWFPANVYKLSPTGSKITPTLPTLPTAQAQVAVCOVIGMD 1158
1045 KNTSGMWGELQARQKROIGWFPANVYKLSPTGSKITPTLPTLPTAQAQVAVCOVIGMD 1097
1159 YTAQNDDELAASKQOIIVLAKEDPMWKGVSQVGLFPSNYVYKLTMDPDSQ 1213
1098 YMANNEDELANSKQOIIVLAKEDPMWKGVSQVGLFPSNYVYKLTMDPDSQ 1152

```

```

RESULT 14
AA71160
ID AA71160 standard; Protein: 1683 AA.
XX
XX AA71160;
XX
XX 08-SEP-2000 (first entry)
XX
DE Rat phosphodiesterase interacting protein, M14.
XX
XX Rat; phosphodiesterase interacting protein, M14; PDE; cAMP-PDE;
XX cyclic adenosine monophosphate phosphodiesterase; antiasthmatic; asthma;
XX antiinflammatory; antipneumonia; dermatological; antibacterial; shock;
XX analgesic; immunosuppressive; antitumor; vasodilator; antihypertensive;
XX antidiabetic; urticaria; antiallergic; antiatherosclerotic; diagnosis;
XX antineoplastic; treatment; inflammatory disease; psoriasis; arthritis;
XX atherosclerosis; Crohn's disease; cystic fibrosis; chronic bronchitis;
XX eosinophilic granuloma; proliferative skin disease; ulcerative colitis;
XX reperfusion injury; atopic dermatitis; diabetes insipidus;
XX conjunctivitis; adult respiratory distress syndrome; allergic rhinitis;
XX arterial restenosis; ankylosing spondylitis; transplant rejection;
XX graft versus host disease.
XX
XX Rattus sp.
XX
XX WO200027661-A1.
XX
XX 18-MAY-2000.
XX
XX 12-NOV-1999; 99WO-US26860.
XX
XX 12-NOV-1998; 98US-0108255.
XX
XX (STRD ) UNIV LEIAND STANFORD JUNIOR.
XX
XX Conti M, Pahlke G;
XX
XX WPI; 2000-376479/32.
XX
XX Polynucleotide encoding a phosphodiesterase (PDE) interacting
XX polypeptide, useful for diagnosis and treatment of asthma, cystic
XX fibrosis, Crohn's disease, and rheumatoid arthritis -
XX
XX
XX Disclosure; Fig 6; 77pp; English.
XX
XX The present sequence is a phosphodiesterase (PDE) interacting protein,
XX M14 from rat. The protein modulates the functions and properties of PDEs,
XX specifically cAMP-PDEs, and also targets them to specific subcellular
XX compartments. The present sequence
XX can be used in the diagnosis and treatment of disease conditions
XX associated with PDE activity. The diseases include asthma, cystic
XX fibrosis, inflammatory airway disease, chronic bronchitis, eosinophilic
XX granuloma, psoriasis, proliferative skin diseases, endotoxemic shock,
XX septic shock, ulcerative colitis, Crohn's disease, reperfusion injury,
XX inflammatory arthritis, atopic dermatitis, urticaria, adult respiratory
XX distress syndrome, diabetes insipidus, allergic rhinitis, allergic
XX conjunctivitis, vernal conjunctivitis, arterial restenosis,
XX atherosclerosis, inflammatory diseases associated with irritation and
XX pain, rheumatoid arthritis, ankylosing spondylitis, transplant
XX rejection and graft versus host disease, disease conditions associated
XX with hypersecretion of gastric acid, and disease conditions in which
XX cytokines are mediators.
XX
XX
XX Sequence 1683 AA;
XX
XX Query Match 48.6%; Score 3046; DB 21; Length 1683;
XX Best Local Similarity 50.8%; Pred. No. 1,1e-186;
XX Matches 652; Conservative 187; Mismatches 267; Indels 178; Gaps 33;
XX
XX 1 MAQFPFPGSLDVAITVEERAKHQDQFLSKPIAGFTIGDQARNFFQSGLPQVLAQ 60
XX 2 MAQFPFPGSLDVAITVEERAKHQDQFLSKPIAGFTIGDQARNFFQSGLPQVLAQ 61

```


Qy	61	IWALADNMNDGRMDQVEFSIAMKLIKLIQGYOLPSTLPVVMKQOPVAISSAPAFGIGI	120
Db	61	IWALADNMNDGRMDQVEFSIAMKLIKLIQGYOLPSTLPVVMKQOPVAISSAPAFGIGI	120
Qy	121	ASMPPLTAVPVPMGSIIPVGMSPPLVSSVPAAVPPLANGAPVIOPLPAFAHPAATWP	180
Db	121	ASMPPLTAVPVPMGSIIPVGMSPPLVSSVPAAVPPLANGAPVIOPLPAFAHPAATWP	180
Qy	181	KSSSFBSGPGSGLNTXKQKASFDVASAPPAEWAHPQSSRLKYROLFNSHDKTMSGHL	240
Db	181	KSSSFBSGPGSGLNTXKQKASFDVASAPPAEWAHPQSSRLKYROLFNSHDKTMSGHL	240
Qy	241	TGPQARTILMOSSLPQAOLASIMNLSIDIDQDKLTAEFFILAMHLIDVAMSGOPLPVLP	300
Db	241	TGPQARTILMOSSLPQAOLASIMNLSIDIDQDKLTAEFFILAMHLIDVAMSGOPLPVLP	300
Qy	301	PEYIPPSFRVRVSSGSGSVISSSVDOQLPEEPSSSEDOQ-EKKLPVTFEDKKRENER	359
Db	301	PEYIPPSFRVRVSSGSGSVISSSVDOQLPEEPSSSEDOQ-EKKLPVTFEDKKRENER	359
Qy	360	GSVELEKROALLEQOKKEQERLAQLEBAEQEKERERQOEPAKQOLEKQLEKQRELE	419
Db	361	GNLELEKROALLEQOKKEQERLAQLEBAEQEKERERQOEPAKQOLEKQLEKQRELE	420
Qy	420	ROREEERREIERREAAKKELEEROROLEWERNRROELNORNKEOEGTVLAKARRKTLF	479
Db	421	ROREEERREIERREAAKKELEEROROLEWERNRROELNORNKEOEGTVLAKARRKTLF	480
Qy	480	ELEALNDKKHGLEGLDIDIRCLATQROEIESTNKSRELIAEITHLQOOLQESQOMLGR	539
Db	481	ELEALNDKKHGLEGLDIDIRCLATQROEIESTNKSRELIAEITHLQOOLQESQOMLGR	540
Qy	540	LIPEKQIISDQIKOYQONSIRHDSILTLKRALBAKELARQOLREQJDEVERETRSKLOEI	599
Db	541	LIPEKQIISDQIKOYQONSIRHDSILTLKRALBAKELARQOLREQJDEVERETRSKLOEI	600
Qy	600	DVFNNQLEKEIREIHSKOOLQOKORSLEAARLKQOEORKSLEL	641
Db	601	DVFNNQLEKEIREIHSKOOLQOKORSLEAARLKQOEORKSLEL	642

Search completed: December 4, 2003, 15:18:34
Job time : 45 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 4, 2003, 15:12:19 ; Search time 24 Seconds
(without alignments)
4860.525 Million cell updates/sec

Title: US-09-674-237A-3
Perfect score: 6269
Sequence: 1 MAQFPTPGSGSLDVAITVEERAKHDQFLSKPTAGFTGDQANFFFGSLPQVLAQ 1213

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 9616862 residues
Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: pir76:*
2: pir1:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5063.5	80.8	1270	2 T09194	adaptor protein in
2	1735	27.7	1011	2 T13055	dynamitin associated
3	1698.5	27.1	1094	2 T13053	dynamitin associated
4	1333	21.3	1097	2 T31504	hypothetical prote
5	506.5	8.1	751	2 T34490	hypothetical prote
6	486	7.8	897	2 A54696	EGF receptor subu
7	485.5	7.7	896	2 S43074	epidermal growth f
8	410.5	6.5	1480	2 S48440	poly(A)-specific r
9	403	6.4	1381	1 S45781	probable calcium-b
10	391.5	6.2	1407	1 S28589	trichomyalin - rab
11	383.5	6.1	1909	2 A45592	liver stage antige
12	381	6.1	1898	1 A45973	trichomyalin - hum
13	375.5	6.0	1549	1 A40691	trichomyalin - she
14	359.5	5.7	737	2 T15597	hypothetical prote
15	359.5	5.7	1017	2 T15598	hypothetical prote
16	342	5.5	2526	2 T20531	hypothetical prote
17	342	5.5	2722	2 T20532	hypothetical prote
18	340	5.4	2738	2 E88320	protein P07A1.6 l
19	331.5	5.3	1069	2 T00377	KIAA0642 protein -
20	330	5.3	793	1 JH0628	caldesmon - human
21	327	5.2	1794	2 T38459	hypothetical diver
22	324.5	5.2	1181	2 C86349	P8K7.4 protein - A
23	323.5	5.2	408	2 T42650	hypothetical prote
24	321	5.1	665	2 JCT191	85K c-Cbl-interact
25	319.5	5.1	1027	2 T46481	hypothetical prote
26	318	5.1	699	2 E84565	hypothetical prote
27	316.5	5.0	816	2 T17257	hypothetical prote
28	314.5	5.0	887	2 G88484	protein F23F12.8 l
29	309	4.9	4574	2 G02520	plectin - human

30	308.5	4.9	1701	2 T09127	probable erythrocy
31	307.5	4.9	1110	2 I51116	NF-180 - sea lamp
32	306.5	4.9	771	1 A33430	h-caldesmon - chic
33	304.5	4.9	585	1 A24168	involucrin - human
34	304.5	4.9	958	2 S47179	hypothetical prote
35	302	4.8	1052	1 A44937	kinetoplast-associ
36	302	4.8	4684	2 A59404	plectin (imported)
37	299	4.8	1233	2 T30989	serine/threonine p
38	297	4.7	2101	2 A42184	nuclear mitotic ap
39	296.5	4.7	3488	2 T34418	hypothetical prote
40	296.5	4.7	4957	2 T03455	ALR protein - huma
41	296.5	4.7	5262	2 T03454	ALR protein - huma
42	296	4.7	4687	1 A39638	plectin - rat
43	295	4.7	729	2 T50989	hypothetical prote
44	294	4.7	746	2 T47237	myosin II heavy ch
45	293	4.7	2442	2 T08621	centrosome associa

ALIGNMENTS

RESULT 1									
T09194 adaptor protein interseectin - African clawed frog									
C:Species: Xenopus laevis (African clawed frog)									
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000									
C:Accession: T09194									
R:Ramadhai, M.; Hoffman, N.G.; Hardison, N.L.; McPherson, P.S.; Castagnoli, L.; Casareni, J. Biol. Chem. 273, 31401-31407, 1998									
A:Title: Intersectin, a novel adaptor protein with two eps15 homology and five src homol									
A:Reference number: Z16605; MUID:99030416; PMID:9813051									
A:Accession: T09194									
A>Status: preliminary; translated from GB/EMBL/DBJ									
A:Molecule type: mRNA									
A:Residues: 1-1270 <YAM>									
A:Cross-references: EMBL:AF032118; NID:g2642624; PIDN:AAC73068.1; PID:g2642625									
A:Experimental source: cell type oocyte									
C:Function:									
A:Description: involved in endocytosis									
C:Keywords: endocytosis									
Query Match									
Best Local Similarity 80.8%; Score 5063.5; DB 2; Length 1270;									
Matches 980; Conservative 98; Mismatches 123; Indels 15; Gaps 8;									
QY	1	MAQFPTPGSGSLDVAITVEERAKHDQFLSKPTAGFTGDQANFFFGSLPQVLAQ	60						
DB	1	MAQFPTPGSGSLDVAITVEERAKHDQFLSKPTAGFTGDQANFFFGSLPQVLAQ	60						
QY	61	ITWALADNMNDNRMOVERSIAMKLIKTLQGYQLPSTIPVWKQOPVAISSA--PAGIG	118						
DB	61	ITWALADNMNDNRMOVERSIAMKLIKTLQGYQLPSTIPVWKQOPVAISSA--PAGIG	118						
QY	119	GIASWPTTAAPVAVVPGMSIPVVGMSPPVSSVPAAPVPLANGAPVTPQIPAPAPHAAT	178						
DB	121	GIVGIPPLAAYAPVPMPIPVVGMSPPLVSSVP--TVPLSNGAPAVYQSHPAFAH-SAT	177						
QY	179	WPKSSSFSGSGPQSLNTKLOKQSFVYASPPAEMAVPOSSRLKQRLFNSHDKTWSG	238						
DB	178	LPKSSSFSGSVAGSQINTKLQKQSPDVAPPLVAVEMAVPSSRLKQRLFNQDKTWSG	237						
QY	239	HLTGPAQTILMOSSLPAQALASITWNLSDIDQDGLTAEFTILMHLIDVAMSGQPLPV	298						
DB	238	NLTGPQARTILMOSSLPAQALASITWNLSDIDQDGLTAEFTILMHLIDVAMSGQPLPV	297						
QY	299	LPPEYIPSPFRVSGSGSVISSSSVDQRLPEPSSSEBQOPKKLPVTEPKKRENF	358						
DB	298	LPPEYIPSPFRVSGSGSLTSSSVSDQRLPEPSEBQOPKKLPVTEPKKRENF	357						
QY	359	RGSVLEKRRQALLBQKKEQERLAQLERABQKKERERQEQAKQLEKQLEKQREL	418						
DB	358	RGNLELEKRRQALLBQKKEQERLAQLERABQKKERERQEQAKQLEKQLEKQREL	417						


```

Db 779 ----- 778
QY 998 EEFIAMTYEESGQDLTFQOQDVIVTKKQDMWTGVDKSGVPPSNVYRLKDSGSG 1057
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 779 EYIYAIAPYESABEGDLSTFSAEMWVVIKKEGEMWITIGSRGTGMPSPSYVQADVGTAS 838
QY 1058 YAG-----KTGSLGKPEIAQVIAVYATGPEQTLAPGQIILIRKKNPGGMEGEL 1110
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 839 TAAAEVPSLIDQGMRAKSEIAQVIAPEATSTEQSLTRGQIMIRKKTDSGMEGEL 898
QY 1111 AGCKRQIOWFPANPKYKLSFG--TSKITPTTELPTKAVQAVNC-QVIGMIDTIAQNDDEL 1167
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 899 AGRRRQIOWFPATYKVIQGGNSGRNTFVSGSRILEMTEQILDKVIALYPPYKAQNDDEL 958
QY 1168 AFSKQIINVLNKEDPDMWKGVSQGVGFPPSNV 1202
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 959 SFDKDDIISVLRGDEBEMWKGELNGSLFPSPNV 993

RESULT 3
T13053
dynamlin associated protein isoform Dap160-1 - fruit fly (Drosophila melanogaster)
C1:Species: Drosophila melanogaster
C1:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C1:Accession: T13053
R:Roos, J./ Kelly, R.B.
J. Biol. Chem. 273, 19108-19119, 1998
A1:Title: Dap160, a neutral-specific Bps1s homology and multiple SH3 domain-containing pro
A1:Reference number: Z17594; MID:98334647; PMID:9668096
A1:Accession: T13053
A1:Status: preliminary; translated from GB/EMBL/DBJ
A1:Molecule type: mRNA
A1:Residues: 1-1094 <R00>
A1:Cross-references: EMBL:AF053957; NID:92984714; PID:92984715; PIDN:AAC39138.1
C1:Genetics:
A1:Gene: Dap160
A1:Cross-references: FlyBase:FBgn0023388

Query Match 27.1%; Score 1698.5; DB 2; Length 1094;
Best Local Similarity 31.6%; Pred. No. 8.2e-62;
Matches 417; Conservative 201; Mismatches 329; Indels 371; Gaps 33;

QY 11 STDVAITVEERAKHQDFSLKPIAGFITGDQARNFFQSGLPQVLAQIWLADMDND 70
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 4 AYDAMAVTFRRLKYQEGFRALQPGVFTGAQAKFFIQSQPLILQIYALADTSD 63
QY 71 GMDQVSEIAMKLIKLQGYQLPSTLPVWKQQPAVISAAPFGIGIASMP--LT 127
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 64 GKNNIFEISIAKLIKLKGMQVPCVLPSTL-----LSSL-----TGDVPSMTFPGSTS 113
QY 128 AVAPVMSGIPIVWGSPLVSSVPPAAVPPPLANGAPVYQPLPAFAPHPATMPKSSFSR 187
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 114 SLSPLD---PLKGI-VPAVAPVPPVAPVPA--VATVISP-PGVSVPSPGPTPTSN--- 162
QY 188 SPSGSLNTKLQAAQFVDVSAAPPALEMWAVPOSSRLKYROLFNSHKTMSGHLLTGQART 247
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 163 --PSNHTSISERAPSIEVSN--QGEWVAQAQKKTQVFRANRRTSGYITGSAQAG 217
QY 248 IIMQSSLPAQULASINWLSIDIDQKLTAEFTLAMHLIDVAMSGQLPPLPEYIPPS 307
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 218 VIVQSLPQVTLAQITLSDIDGRLNCDDEFILAMFLCEKAMAGSEKIPVTLFQEWVFN 277
QY 308 FFRVRSQSGMSVSSSVQRLPEEPPSD-----EQGEKLP--VTREDK 353
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 278 LKTKISRPG---SVSGVSPSPGSPASRHSVSSQSGVGVADADTAGLPQGTSPEDKR 333
QY 354 RENFERGSVELERQALTEQQRQERLAQLERAEQERERERQEQAKROLTEKOLE 413
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 334 KENYVGAQALDPRKRIMEDQQRKEBERKERERADREKARLEAKRQDELERQ 393
QY 414 KQRELRQREBERKEIREREAARLEERQOLEWERNRROBELLNQRNKEQEGTVLAKR 473
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 394 RQREIMERQQRKRELAKARKELEKQROQEWEGQATIAEWNAQKEREQERVLKQAK 453

```

```

QY 474 RKTLEFEELANDKKHOLEGKLQDIRCRLATQROEIESTNKSRELIATITLQOOLQES 533
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 454 NQULVELSTLNKELKELSRICDTRAGVTANKYVTDGMRTRDRDSMSMSQKARIKQ 513
QY 534 QOMLGRLLPEK-----QILSDQLQVQNSLRHSDLLTLKRALEAKELARQQLR 582
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 514 NAKLLQTLQERAKWEAKSKASGALGEMVAQOEQ-----LNAFAHKLIIINQJK 563
QY 583 EQLDEVERETRKLOEIDVFNNQLKELE-----IHSKQLOKQSLSEAKRKQKEQEK 637
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 564 DKVENISKEIESKEDINNDVQMSLEKALSALTTCEDILKEVDVQRTSVLELKYNRK 623
QY 638 SLELEKQEDAQRVOERDKQWLEHVOEEQPRPKPHEDRLTKEDSVYKKEAERAPK 697
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 624 NETSVSSAMDT-----GSSSNM-----BETGTYTDPY-----AVASDISALAP 664
QY 698 EMQDKQSRLEFHPHOEPAKLATQAPWSTTEKGPLTISAQESVYVYRYALPPERSHDEI 757
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 665 AYD-----LQGPAP-----EG-----FVKYQAVYEFNANNAEEL 693
QY 758 TIQPGDIWVWDESQTGEFGWLGELKGTGMPFANYAEKIPENEPPTPAKPVTLDTLTA 817
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 694 TEVPEDIIIVPLEQNAEFGMLAGEINGHTGMPESYVEKLVEGEV-----A 739
QY 818 PLTALREPPAPLPVTSSESTTPNNMADPSSSTWSPSSNEKPEPTDMMDTMAQPSLTVPSA 877
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 740 PFAAA-EPAPVDAQVADT-----YNDININ-----SSTIPA 768
QY 878 GOLRQSAFTPATATGSSPSPTLQGEKEVEGLQAQALYPMRAKKNHNLFNKSDVITVLE 937
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 769 S-----ADLTAG-----DY----- 778
QY 938 QQDMMWFGEVQQCKMFPKSYVKLISGPVKSTSIDTPTSPASLKRVASPAKAPALPG 997
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 779 ----- 778
QY 998 EEFIAMTYEESGQDLTFQOQDVIVTKKQDMWTGTVGDSGVPPSNVYRLKD----- 1052
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 779 EYIYAIAPYESABEGDLSTFSAEMWVVIKKEGEMWITIGSRGTGMPSPSYVQADVGTAS 838
QY 1053 -----SESG----- 1057
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 839 TAAAEVPSLIDQETLNGAATAPAEQEVQPLPVQEPSEBPISPGAGABAHND 898
QY 1058 -----TAGKT-----GSLGKPEIAQVIAVYATGPEQTL 1087
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 899 IDTEVSQINTQSKTOSSEPAESYSRPMGRTSSMTPGMAKRSIEIAQVIAPEATSTEQLS 958
QY 1088 LAPGQIILIRKKNPGGMEGELQARGKKRQIOWFPANVYKLSFG--TSKITPTTELPTKA 1145
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 959 LTRGQIMIRKKTDSGMEGELQARGRRQIOWFPATYKVIQGGNSGRNTFVSGSRLE 1018
QY 1146 VQPAVC-QVIGMYDTAQNDELAFSAKQIINVLNKEDPDMWKGVSQGVGFPPSNV 1202
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1019 MTEQILDKVIALYPPYKAQNDDELSTFDKDDIISVLRGDEBEMWKGELNGSLFPSPNV 1076

RESULT 4
T13104
hypothetical protein Y116A8C.36 - Caenorhabditis elegans
C1:Species: Caenorhabditis elegans
C1:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C1:Accession: T13104
R:McMurray, A.
submitted to the EMBL Data Library, October 1999
A1:Accession: T13104
A1:Status: preliminary; translated from GB/EMBL/DBJ
A1:Molecule type: DNA
A1:Residues: 1-1097 <W11>
A1:Cross-references: EMBL:AL117204; PIDN:CAB55138.1; CESP:Y116A8C.36
A1:Experimental source: clone Y116A8C

```

C:Genetics:

A:Gene: CESP:X116A8C.36
A:Introns: 50/3; 245/3; 411/3; 486/3; 697/1; 820/3; 926/2

Query Match 21.3%; Score 1333; DB 2; Length 1097;

Best Local Similarity 29.4%; Pred. No. 5,6e-47;
Matches 369; Conservative 200; Mismatches 454; Indels 234; Gaps 39;

```

15 WAITVEERAKHDQQLSLKPIAGFTTQDQ-----ANPFQSGLPQPVLAQIWMALDM 67
Db WEVSAAEQKPFAMF-----GQLTGQPFMDAVTANALMRSNLPQVLSQIWMALSD 57
5 NNDGMDQVSIAMKLIKLGQQLPSTLPVVKQOPVAISSPAGIGIASMPPLT 127
68 NNDGMDQVSIAMKLIKLGQQLPSTLPVVKQOPVAISSPAGIGIASMPPLT 127
58 DKDGLDREVSIMRLALNCLAGIPPIPPQLSLVPA--RANPPTWPS----- 107
128 AVAPVPMGSIPIVNGSPPVY-----SSVPAVAVPLNAGAPVYIOPLPAPAPAAATWPK 181
108 -----RHGSVDYSQTLPPALIDRRMSQYIPSA--FVSIAGTP----- 142
182 SSSFSRSGSGSOLNTKQKAGSPDVASAPPAEAMAVPOSSRLKTRQLFNSHDKTMSGLT 241
143 SSRNHSISAGSPLN-----NDRNVEGRQLENNWAIIPHNNKLKYSQLFNALDKERTGLS 196
242 GPQARTIMOSLSPOAOLASTIWNLSDDODGLTAEPFLAMHLIDVAMSGQPLPVLP 301
197 SQVGRSALGSLPNTVLAHITFLSDVNDGKLSVDEVAISQYMLEMFKSGFALPKITPL 256
302 EYIPSPFRVRSGSGMSVSSSVQDRLPEEPSSEDEQOEKCLPVTEDEKRENFERGS 361
257 ELV-----RMCGISSRANNTPELEBGAEPPO--KSPAKTFEDKQDVLKSGQ 303
362 VELEKROALLBQKKEQERLAQLERAQERKERQOEAKQOLEKOLEKOLEKOLEK 421
304 AELERRRRVLEEEERRRAVEKKEERBEAKKNRQEKERQAEVERQAELEERQIIIAQ 363
422 REEE---RRKEIERR-----EAKKELE---ROQLMEMNRROELNORKEQEGTVV 469
364 REEEKRRLEMERREDEDEKRRKQMEKAKKQOVQMEPKKNPYNQKQOEHERLAQ 423
470 LKARRKLEFLEALNDKQKLEGKLODIRCLATQROEIBSTNSRELIATIEITHLQOQ 529
424 RQQRKTLQFQALDEKVIDEVDIGAKAEVAEVGTGFIEMRSTREKVARL-----KE 479
530 LOESQOMGLRILPEKQIISDQKQVQNSLRHDSLITLKALEAKELARQOLREOLD-- 586
480 LOETNQ---KTAIESQELGHQV--LOKQSAHNET--TORKSELEALRRKRDRIKKAIEDPA 533
587 -EVERETSQKQEIIVF--NNOLKELREIHSKQQLQKQSLBAARLKQOEKESLEKQ 644
534 LELSTEKESKYNQTEILKTNKEKYTDVYSK-----LVAKKEEYRNSFEL-- 578
645 KEDAQRVQERDKQMLEHVQOEOPRPKP-----HEEDRLKREDSVRK 688
579 ----LVHAQTHARSKIGFEPAKASAPASAPAPAPATTNNGPANNDAFGED--K 631
669 KEAERARAPENMODKQSLRPHRQEPAKLATAQPMSTTEKGPITISAGSVKVVYRALYR 748
632 TDAQSRFPADFGATST-----ADPPAQIACAP--AHSKGAVDQSAFNIHDTYKCRALFA 683
749 FESRSHDELTTOPGDIWVWDSQGEPMGLGELGKGTGMPANVAEKIPENEVTPAKP 808
664 FEAREDELSFEPGVITVFOSHAAPGMRAGQLEKQGMPEAVEAIA--AVPTPG-- 739
809 VTDLTSAPAKLALRETPAPLPVTSSEPTTPNNNADSSSTWSSSNKEPETDNMDTAA 868
740 -----GDPLIQMPPNMTPESSVDQIGV--KAARKAEI-----AA 772
869 QPSLTVPSAGQLQRSAFTPATATSSSPSPVLGCGEKEVGLQALYIMPRAKGNHILFN 928
773 AMGLTEGA-----PPASSAPAAAAVIS-----QCIAGFQMRANNEDELSFA 814
929 KSDVITVLEQDMMWFG--EVQGGKGMFPKSYKILSGFVRKSTSLDTGPTESPAKLKVA 987

```

```

Db 815 KGDITVELEKQEMKMKRNAPAGEIGWFKSVYKVGATTTSTTPI--VSPKASAGAPGAA 873
Qy 988 SPAKPAIPGEE-----FTAMTYSSBEGDLTPQOGDVIYVTKKQDWMWG 1034
Db 874 AGAAYDVAPSVTLQASETAPQOQLYVIVDFEAVETTDIALAHGDTITLVLEKDEWKG 933
Qy 1035 TVGDGSGVPPSNVYRLDSESGTAGKT--GSLCKKEI---AQVIASVATGPEQTLA 1089
Db 934 RGNREGIFPANYVEISVQAGDPTPTQAPTPAPPTVLCBAKVVDPAASAPNQGIG 993
Qy 1090 PGQILIRKKNPGMWGEQIARCKKROIGFPAHYVYKLSPTGSKITPTLELPTRAVQPA 1149
Db 994 VGEIVKIRKESAAQMWGELLIRNGK--IAGMFPPEYKVLLEAASPAT----- 1040
Qy 1150 VCQVIGMYDTYQNDDELAFSKQIIVLANKEDPMWK--EVSGQVGLFPSTNYK 1203
Db 1041 --RATAVYDYEASQPDLEIGFTGDVITVTDKSEAMWMSGHREDDPSKGLFPSTNYQ 1095

```

RESULT 5

T34490
hypochemical protein ZK1248.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999

C:Accession: T34490

R:latreille, P.
submitted to the EMBL Data Library, June 1995

A:Description: The sequence of C. elegans cosmid ZK1248.

A:Reference number: 221534

A:Accession: T34490

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-751 <LAT>

A:Cross-references: EMBL:U02244; PIDN:IACT1084.1; GSPDB:GN00020; CESP:ZK1248.3

A:Experimental source: strain Bristol N2; clone ZK1248

C:Genetics:

A:Gene: CESP:ZK1248.3

A:Map position: 2

A:Introns: 37/1; 74/3; 118/3; 355/2; 661/1; 728/2

Query Match 8.1%; Score 506.5; DB 2; Length 751;
Best Local Similarity 24.7%; Pred. No. 1.3e-13;
Matches 214; Conservative 115; Mismatches 273; Indels 263; Gaps 34;

```

15 WAITVEERAKHDQQLSLKPIAGFTTGDQANFFQSGLPQVLAQIWMALDNNDGRMD 74
Db 111 WPIPTQAKYDSIFQSLNPVNGKLSGAHYRPVLMNSGLDAHALARIWEISDDQKDNLD 170
Qy 75 QVESIMAMKLIKLGQQLPSTLP--VMKQOPVAISSAPAGIGIASMPPLTAVAPV 132
Db 171 RIEMSVALLHYRSLQSDPVPAPQLPVNLHPSKMYAHSSNF-----AAPP--HPPR 221
177 -ATPDKSSFSRSGSGSOLNTKQKAGSPDVASAPPAEAMAVPOSSRLKTRQLFNSHDKT 235
222 PMGSSRAGSVTSLDDVAMSSQ--YSATMPRAVP-----QGRAYSAQAHNGSRIS 271
177 -ATPDKSSFSRSGSGSOLNTKQKAGSPDVASAPPAEAMAVPOSSRLKTRQLFNSHDKT 235
272 GASTPISASHS-----IHSF-----PAGEWPIITNGD--YADQPAQTDTN 308
236 MSGHLTPQARTIMOSLSPOAOLASTIWNLSDDODGLTAEEFLIMHLIDVAMSGQPL 295
309 KDGLVDMQDMKAPMTTGLSAQILAHWALADIKKQQLMBQFALTMHLIDMAKRGESI 368
296 PPLVPPPEYIPSPFRVRSGSGMSVSSSVQDRLPEEPSSEDEQOEKCLPVTEDEKRE 355
369 PSELPLHLIPSPFRPPEPRLHHPAQSVESTPQLEPASTWEIKALE-----GE 417
356 NFERGSELEKRRQALLBQKKEQERLAQLERAQERKERQOEAKQOLEKOLEKOLEK 415
418 NEF--MKQLAESIQSMVLERTAEAVIQLE--ADMTIKNSSIK-----NLQWELATLESTV 470

```

```

QY 416 RELEROREBERK-----EIERREAAKELEROROLEWERNROELLNQRNKEOGTVV 469
DB 471 KQLEKQKGAETRLADVDYQIOLESB-----CAQKQETKEDTERKMQOIDEBAKNED--- 524
QY 470 LKARKRTLEFELBALNDK--HOLEBKLODIRCLATOROEIESTNKSRELIAETHTQ 527
DB 525 CKA-----NDEKEMELKEIEMLDNQFVTGGEIVKETSQGEQVAELT--- 569
QY 528 QOLOESQOMLGRILPEKQILSDOLKQVOQNSLHRSDLLTKRALENKELARQOLR-EQID 586
DB 570 -----TLERKE-ARDQIOHEKID 586
QY 587 -EVERETSKLOEIDVFNQNLKEIREHSGKQOLQKORSLEAARKQKEQERSLELEKQ 645
DB 587 AAIEMTTKLTQVSDAVESKEEMQOI-----LRSGQLSLTVIDQSLSDTVGCTG 641
QY 646 EDAQRVRQERDKOMLEHVQOESQPRPRKPEEDRLKREDSVRKKEAEERAKEMODKOR 705
DB 642 TSSQ-----NHVQQPPDP-----FASAPANPAAD----- 665
QY 706 LFHPHOEPKATQAAPWSTTEKGPLTISQESVKVYVYALYPFESRSHDEITIQGDIV 765
DB 666 -----PFQVQD-----FGSSGH----- 678
QY 766 MVDESQTEGPGWLGELKGTGMPPAN--YAEKIPENEVPTPAKPYTDLTSAAPAKLALR 823
DB 679 -FDAFPPTDPFAQGG-----FSDSGFAQSAAPK--PAAP-----RPAPEKAKR 719
QY 824 ETPAPLPVTSS-PSTTPNNMADFS 847
DB 720 ETPVNDPFAPSOQOSTOPAGPADFA 744

```

```

RESULT 6
A:Accession: A54696
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-897 <FAZ>
A:Cross-references: GB:121768; NID:G404756; PIDN:AAA02912.1; PID:G404757
C:Superfamily: calmodulin repeat homology
C:Keywords: EF hand; phosphoprotein
F:48-80/Domain: calmodulin repeat homology <EF1>
F:160-192/Domain: calmodulin repeat homology <EF2>
F:223-255/Domain: calmodulin repeat homology <EF3>

```

```

Query Match 7.8% Score 486; DB 2; Length 897;
Best Local Similarity 21.3% Pred No.1e-12;
Matches 205; Conservative 138; Mismatches 298; Indels 322; Gaps 31;
QY 15 WAITVEERAKHQDOQLSLKPIAGFTTGDQARNFFQSGLPQVLAQIYALADNMNDGRMD 74
DB 122 MAVKSEDRKAKYDAIFDSLSFPVDGFLSGDKVKFVLNSKLPVEILGRVWELSDIDHDGKLD 181
QY 75 QVEFSIAMKLKIKLQGYLPESTLPVWKQOPVALISSAPAFIGGIASMPPLTAAVAVPM 134
DB 182 RDEFAVAMEFLVYCALE-----KEPYVM 203
QY 135 GSIPVVGMSPLVSSVYPAAPVPLANGAPVPIQPLPAFAHPATWPKSSFSRSGPQSL 194
DB 204 -----SLPEPLVPP-----SKR 215
QY 195 NTKLOKQAGFDVASAPPAEMAVPOSSRLKTYRQLFNSHDKTMSGHLTGPQARATILMQSSL 254
DB 216 KT-----WVVSPEAKAKYDEIFLTKTDKMDGYYSGLEVAETFLKCTL 257

```

```

QY 255 PQAQLASTWNLSIDIDQDKLTAEERILAMHLIDVAM-SGQPLPVLPPEYTPPSRRVRS 313
DB 258 PSALIAHWLSLCTDYKCGCKLSKQFALFHLINQKLIGIDPPHSLTPEMPPS---DRS 314
QY 314 GSGMVISSSSVQDQLPEEPSSEDEQPEKKLPVFPEDEKKEENPFGSVLEKRRQALIE 373
DB 315 SIQKNITGSSPV-----ADFSAIK-----ELDTLNNELVD 344
QY 374 QQRKEORLAOLERAQERKERERQOBAKQOLELEKOLEKORLELORREBERREIERR 433
DB 345 LQREK-----NNEVQLKEKE-----DTVQRTSEVQ---DLQDEVQRE 380
QY 434 EAKKELEROROLEWERNRROELLNQRNKEOGTVVLAARKRTLEFELBALNDKQOLRG 493
DB 381 SINLOKLAQKQ-----QVQELLGE-----LDEQAKOLEE 410
QY 494 KLQDIRCLATOROEIESTNKSRELIAETHTLOOQLOESQOMLGRILPEKQILSDOLQ 553
DB 411 QLOEVRKCKAEAOILSSLK-----AEITSQESQISSYEBELK-----AREELSR 456
QY 554 VQONSILHDSLLTKRALENKELARQOLREQLDEVERE-----TRSKLOEIDVFNQOLKE 608
DB 457 LQOETAQ-----LEESVSGKAQLEPLQOHLQESQOERISSWQRLKEMKDLTDNNQ--- 507
QY 609 LREHSGKQOLQKQ-----RSLAARLQKQEKRSLELEKQKEDAKQRYQ 653
DB 508 -----SNWSSPQSVLVNGATDYCSLSTSSSETNPFNHAEGQNNLJESBPTHQESSVSS 562
QY 654 ER-----DKOMLEHVQOERQPRPRKPEEDRLKREDSVRKKEA-----EERAKEMQDK 702
DB 563 PEIAPSDVTDESEANTVAGNEKVTFR--FDDDKHKEEDPRVNESSSLTDAVADTNLDF 620
QY 703 QSRLE-----HPQEPKATQAAPWSTTEKGPLTISQESVKVYVY 743
DB 621 QSDPFGVSDPFKMDPFKIDPFQGPFPFGSPDFADCFKQSTDPFTTSSSTDPSASNN 680
QY 744 RALYFESRSHDETIOGDIYMWDESQTEG--GWLGEKLGKGTGMPPANYAEKIPENE 801
DB 681 SNTSVEETWKHNDPPAPGTVVAAASDASATDPFASVFPNESRG-DGFADFSTLSKVNND 739
QY 802 VTP-----AKPVTDLTSA-----PA--PKLARETPAPLPVTSEPTTNN 842
DB 740 AFNPTISSSTSSVTIAKMLEETASKSDVPALPPKGTPTPPCPPP-----PKRRPLN 794
QY 843 WADFSST-----WPSSNKEPFTDNDMTAAOPLSTVPSAGOLRORSAPFPATATGSS 895
DB 795 KLDSSDPLKLNDDPFQPPFGNDSPEKXKDPDMC-----DPTFSSTTKKE 838
QY 896 PSP 898
DB 839 ADP 841

```

```

RESULT 7
S43074
epidermal growth factor receptor substrate - human
C:Species: Homo sapiens (man)
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 01-Dec-2000
A:Accession: S43074; I38525
R:Bernard, O.A.; Muchauffe, M.; Mecucci, C.; van den Berghe, H.; Berger, R.
Oncogene 9, 1039-1045, 1994
A:Title: A novel gene, AF-1p, fused to HRX in t(1;11)(p32;q23), is not related to AF-4,
A:Reference number: S43074; MUID:94181254; PMID:8134107
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-896 <BER>
A:Cross-references: EMBL:Z29064; NID:G470034; PIDN:CAA82305.1; PID:G470035
R:Wong, W.T.; Kraus, M.H.; Carlomagno, F.; Zelano, A.; Druck, T.; Croce, C.M.; Huebner,
Oncogene 9, 1591-1597, 1994
A:Title: The human ep15 gene, encoding a tyrosine kinase substrate, is conserved in evo
A:Reference number: I38525; MUID:94239734; PMID:8183552

```

A:Accession: 138525
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-821, 'M', 823-896 <RES>
 A:Cross-references: EMBL:U07707; NID:G466259; PIDN:AAA52101.1; PID:G466260
 C:Genetics:
 A:Gene: GDB:EP515; AF-1P; MLT15
 A:Cross-references: GDB:360337; OMIM:600051
 A:Map position: 1p32-1p32

Query Match 7.7%; Score 485.5; DB 2; Length 896;
 Best Local Similarity 20.5%; Pred. No. 1.1e-12;
 Matches 198; Conservative 150; Mismatches 273; Indels 345; Gaps 29;

```

Oy 15 WAITVEEAKKHDOQLSLKPTAGFTTGQARNFPGSGLPQPVLAQIATLADNMNDGMD 74
Db 122 WAVEKEDAKYDAIFDSLPVGVGSGKVKVLLNSKLPVILICRWELSDIDHDGMLD 181
Oy 75 QVEFSIAKMLIKLKQYQLPSTLPVWKQGFVAISSAPAFQIGIASMPPLTAVAPVM 134
Db 182 RDEPVAMFLVYCALE-----KEPYM 203
Oy 135 GSIPVGVMSPLVSSVPPAAVPLANGAPVLIQPLPAHPAATWPKSSFSRSGSQSL 194
Db 204 -----SLPPALVPP-----SKR 215
Oy 195 NTKLOKASPDVVASAPPAEAMVPOSSRLKTRQLNSHDKTMSGHLTGPOARTIMOSL 254
Db 216 KT-----WVSSPAEKAKYDEIFLTKDKMDGVSGLVRELFLLKTKGL 257
Oy 255 POAQLASIMNSLSDIDODGKLTAEFETLMLHLIDVAM--SGOPLPVLPPEYIPSPRRYRS 313
Db 258 PSTLHAHMSLCDTDCCKLSDQKLPALAFHLISQILGIDPHVLTETEMPPS---DRA 314
Oy 314 GSGMSVSSSVYDQRLPEPSSSEDEQPEKKLPVTEDEKRENERGVELEKRRQALLE 373
Db 315 SLQKNIIGSSPY-----ADFSAIK-----ELDTLNNETVD 344
Oy 374 QQRKEQELQALERAEQEREREREOAKQLELEKQLEKQRELERQREERREIRRR 433
Db 345 LQREK-----NNVEQDLKEKE-----DTIKQRTSEVQ---DLQDEVGRE 380
Oy 434 EAAKELEROROLEWERNRROELLNQRNKEGEGTVLAKARKTLEFLEALNDKQHGEG 493
Db 381 NTNLOKLAQKQ-----QVQELDE-----LDEQAKQJEE 410
Oy 494 KLQDIRCLATQROFIESTNKSREIRIAETILHQOQLQESQOMLGRLLPEKQILSDQKQ 553
Db 411 QLKEVRKKCAEBAQLISSLK-----ALTSQESQISTYEBELAK-----AREELSR 456
Oy 554 VQGNLHLDLSLLTRALEAKELAQQLREQLDEVERETR-----KLQELDVFNQGLK 607
Db 457 LQOETAE-----LEESVESGAQLEPLQOHLQDSQOQETSSQMKLMEMKDLNENISQLN 510
Oy 608 ELREIHS-----KOQLKQSRLEAARLKQKEQERKSELE-----641
Db 511 WCSSHSILVNGATDYCSLSSSETANLENHEGQSNLESPIPIQESPARASPELLRSG 570
Oy 642 EKQKEDARRVQERDKOMLEHVQOEOPRPRKPHEDRLKREDSVRKKEAEER-----694
Db 571 VTDEHEVTAAVTEKVCSELDN-----NRHSKEEDPNNVSSSLVGPVADTNLDPFQS 622
Oy 695 -----AKEMQDKQRLRPHQEPK-----LATQAPWSTTEKGPITISAQESV 738
Db 623 DPFVSGDPFKODPFEKIDPFPGDPFKGSDPFASDCFFRQSTDPFATSSSTDPFSANNSI 682
Oy 739 KVVYVRYALYPFESRSHDEITIQGDIIVWVDESQT-----GEPMWLG-----780
Db 683 TSV-----ETLKNDPFAPGQGVVAASDSATPPASVFNESFGGQFADFTLSKYN 734
Oy 781 ---ELKGTGWFPAN-----YAEKIPENEVPTPAKPVTDLTSAAPALALRETPAPLP 830
Db 725 NEDPFRSATSSSVSVVITKNVFEETSVSSEDEPALP-----PRIGPTPRCPPLP 785

```

Oy 831 -----VTSSEB-----STPNWMADESSRWSSNKEPETDNMDT 865
 Db 786 PGRKSINKLSPDFEKLNDPPQPPGNDSPKEXKDEIFCDBFTSATTTINKADPSNPNAN 845
 Oy 866 WAAQPS 871
 Db 846 FSAVPS 851

RESULT 8

S48440 poly(A)-specific ribonuclease (EC 3.1.13.4) - Yeast (Saccharomyces cerevisiae)

N:Alternate names: protein Y1B6; protein Y1R006c

C:Species: Saccharomyces cerevisiae

C:Date: 02-Dec-1994 #sequence_revision 02-Dec-1994 #text_change 21-Jul-2000

C:Accession: S48440; S30889; S58706; S50795; S27443

R:Badcock, K.; Churcher, C.

A:Submitted to the EMBL Data Library, August 1994

A:Accession: S48440

A:Molecule type: DNA

A:Residues: 1-1480 <RAD>

A:Cross-references: GB:Z47047; EMBL:Z38062; NID:G603997; PID:G763351; MIPS:Y1R006c

R:Sachs, A.B.; Deardorff, J.A.

Cell 70, 961-973, 1992

A:Title: Translation initiation requires the PAB-dependent poly(A) ribonuclease in yeast

A:Reference number: S30889; MUID:92405166; PMID:1339314

A:Accession: S30889

A:Molecule type: DNA

A:Residues: 1-234, 'T', 236-265, 'YCCPSGSKN', 274-479, 494-652, 658-1290, 'R', 1292-1395, 'EAFCLH

A:Cross-references: EMBL:M96688; NID:G172095; PIDN:AAA34841.1; PID:G172096

R:Yoss, H.

submitted to the EMBL Data Library, June 1994

A:Reference number: S58703

A:Accession: S58706

A:Molecule type: DNA

A:Residues: 1-1480 <YOS>

A:Cross-references: EMBL:X79743

R:Yoss, H.; Tamames, J.; Teodoru, C.; Valencia, A.; Sengen, C.; Wiemann, S.; Schwager, C.

Yeast 11, 61-78, 1995

A:Title: Nucleotide sequence and analysis of the centromeric region of yeast chromosome 1

A:Reference number: S50795; MUID:95282515; PMID:7762303

A:Accession: S50795

A:Status: nucleic acid sequence not shown

A:Molecule type: DNA

A:Residues: 306-355, 640-688 <VOM>

A:Cross-references: EMBL:X79743

C:Genetics:

A:Gene: SGD:PAU1; MDP3; MIPS

A:Cross-references: SGD:S0001445; MIPS:Y1R006c

A:Map position: 9R

C:Keywords: hydrolase

Query Match 6.5%; Score 410.5; DB 2; Length 1480;
 Best Local Similarity 20.5%; Pred. No. 2.1e-09;
 Matches 261; Conservative 174; Mismatches 444; Indels 395; Gaps 47;

```

Oy 17 ITVERAKHQOQLSLKPT-AGFTGQARNFFQSGLPQPVLAQIATLADNMNDGRMDQ 75
Db 266 ITAQDQAKFETLFRSIVTNGSNTVSGANCRILMRSGLPQQLARIMTLCDTSSKAGELLF 325
Oy 76 VEFISIAKMLIKLKQYQLPSTLPVWKQ-----104
Db 326 PEFALAMHLINDVAGGDTIPYELDSKTKNEVSSFLDAINISIANQDSSANDAPKTPDEF 385
Oy 105 -----QVVAISSAPAFGI-----GJIASM-----123
Db 386 ITAGVQMLQPPPTGVMQTSRGILPQSQITGGVASALNPOSTGFMATTTNMSMNTCTP 445
Oy 124 ---PPLTAVAVPM-----GSIP-VVGSPPVSSVPPAAV-----PPLAN 160
Db 446 GLNFOITGAGASMQPNITGNALQPTTGMMPTTGMMPTTGMMPTTGMMPTTGMMPTTG 505

```



```

QY 161 GAPVIO-----LPAPAPATWPKSSFSRSGPSQ-----194
Db 506 GA-----LQSYTGCGYGVWPOQSPASM--PNLSFNOQIOLOTLGQPOPTGFLPPSNFS 560
QY 195 -----NTKLOKQSFVVASAPPAEWAPOSSRLRYOLFNSHDKTMSGHUL 241
Db 561 ATPLPLAQKTGFQNNNEIYTKSNFNNNLINSSQDKISTEKSFLFYKIFETPTQNGKGLD 620
QY 242 GQARTIMQSSLPQAOIASIWNLSIDDDGKLTAEFFLAILIVAMSGOPLPVLP 301
Db 621 SPTAVEIFRKSGLINRADLEQIWNLCIINNTOQLNKQEFALGMHLYVGLNGKRIPIVLP 680
QY 302 EYIP-----PSF-----RRVSG--314
Db 681 SLIPSTKLLDNKQKTEPTTTKPSFGKIDALSTYKNNDDVLPTNRNRYKVSAN 740
QY 315 ---SGMSVSISSVDORLPPESSSEDEQPEKKLPVTFEDK---KRENFERSVLEKR 367
Db 741 EEOSSFSFSAKSVNH-----SSSTLQTDIDISVDKTEVKTKAKPYAGFSR---EINK 791
QY 368 RQALLQCKRQER-----LAQLEAQRERERQOEAKRQLE 407
Db 792 NIASLENEIKNISNPNCYDSSIPSDLTSPFALIKLNLNLEISTINNEITNAXIOLY 851
QY 408 LEKQLEKQELERQREERKEIERREAKRELERQLEWERNRQELINQNKQEGT 467
Db 852 RKK---NPSSTIGSGNGEITENDRKAKSRALLR-----ABMSALTGSTSEDS 899
QY 468 VYLKARKKLTLEF-LEALNDKHKQLEGLQDIRCRLATOROESTNKRRLAIEITL 526
Db 900 LEMBEDQSAEIKRIQENGKQEI--IKDIRSSISDISAKSMTGSM---ISN- 951
QY 527 QOOLQESQMLGRLLPEKQILSDLKQVQNSIHRDS-----563
Db 952 -QEFERMEFGIGLEDGVRFL--DDLKSNKNSVTBESSPVPSSTPPVDRSSPSYSQF 1009
QY 564 ILTLKALEAKELARQLEOLDEVERETRSKQLEIDVNV--NQLKELAEIHSKQLE 618
Db 1010 KTAEBBAVYLKQAKKQMEKLAKEPKNRNVTQSSRSISSENSRQPOQIAGSSNLVLP 1069
QY 619 -----QKORSLBAARLKQERKSLLEKKE-----646
Db 1070 RATPQEKYVEVAQPTQPVQSTQPVQPTQPVQPTQPVQPTQPVQPTQPVQPTQPVQ 1129
QY 647 DAQRVQERDKQMLEHVQOEOPRPKXPHEDRLK-----EDSVKKEAE-693
Db 1130 NAKQESDEDEDEDEKRLQEBELKRLKLLKKADKEKRLALRKQIEDAQNESDEEETNGK 1189
QY 694 -----RAKP-----EMQDKOSLFFPHQEP--KLATQAPMSTTEKGLPITIA 734
Db 1190 NFGHVNVPQAPVAPSAAPSQNSTNAPSVHAAVTPAAGKSTGIPSTTMGHNPFKCA 1249
QY 735 QESVKKVYRALYFESRSHD-EITIQPDIVWDESQGEPCMLGELKGTGMPPANY 793
Db 1250 SAS-----STSTPARAAEMQRKQRG-----LDDE--DQMSDEDSNNR-----VAV 1292
QY 794 AEKIPENEVPTP---AKPVTDLTAPAPKALRETAP--LPVTSSEPTTPNNMADS 847
Db 1293 DNKVEBAKIGHPDHAPAPV---TAAPLPV---TPVPAPVVPQANTSNEKSSPIPIA 1345
QY 848 STWPSSNS-----KPTDNMDTMAQPSLTVPASQALQRS--APAPATATGSSPPVVG 901
Db 1346 PIPSTVQEPVPVLPAPLVADQFQEPPIPSAPALATAVOKSGSSTPALLAGVLP--1402
QY 902 QOEKVEGLQAOALYPMRAKKNHLENK-----SDVITVLEEQDMMWFGEVQOK 951
Db 1403 --PILPTQOASTSEPIIAHVNVNGAEKGTGAYGSDSDDIVLSIPS-----VGTDEEE 1455
QY 952 GMPFKSYVKLISGP 965
Db 1456 GAQPVSTAGIPSTP 1469

```

```

RESULT 9
S45781
probable calcium-binding protein YBI047c - yeast (Saccharomyces cerevisiae)
N:Alternative names: hypothetical protein YBI0520
C:Species: Saccharomyces cerevisiae
C:Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text change 19-Apr-2002
C:Accession: S45781; S50284; S45782; S39841; S37339; S42498
R:Goffeau, A.; Joniaux, J.L.; Purnelle, B.; Skala, J.; de Wergifosse, P.; van Dyck, L.
submitted to the Protein Sequence Database, August 1994
A:Reference number: S45745
A:Accession: S45781
A:Molecule type: DNA
A:Residues: 1-961 <GDP>
A:Cross-references: EMBL:Z35808; GSPDB:GN00002; MIPS:YBI047c
A:Experimental source: strain S288C
R:de Wergifosse, P.; Jacques, B.; Joniaux, J.L.; Purnelle, B.; Skala, J.; Goffeau, A.
Yeast 10, 1489-1496, 1994
A:Title: The sequence of a 22.4 kb DNA fragment from the left arm of yeast chromosome II
NA-binding protein.
A:Reference number: S50284; MUID:95176707; PMID:7871888
A:Accession: S50284
A:Molecule type: DNA
A:Residues: 1-961 <DEP>
A:Cross-references: EMBL:X78214
A:Experimental source: strain S288C
R:Dubois, E.; El Bakoury, M.; Giansdorff, N.; Messenguy, F.; Pierard, A.; Scherens, B.,
submitted to the Protein Sequence Database, August 1994
A:Reference number: S45782
A:Accession: S45782
A:Molecule type: DNA
A:Residues: 579-1381 <DUB>
A:Cross-references: EMBL:Z35808; GSPDB:GN00002; MIPS:YBI047c
A:Experimental source: strain S288C
R:Scherens, B.; el Bakoury, M.; Vlerendeels, F.; Dubois, E.; Messenguy, F.
Yeast 9, 1355-1371, 1993
A:Title: Sequencing and functional analysis of a 32 560 bp segment on the left arm of ye
A:Reference number: S39824; MUID:94205266; PMID:8154187
A:Accession: S39841
A:Molecule type: DNA
A:Residues: 579-1381 <SCH>
A:Cross-references: EMBL:Z23261; NID:9313733; PIDN:CAAB0797.1; PID:9313748
A:Experimental source: strain S288C
C:Genetics: SGD:EDI; MIPS:YBI047c
A:Gene: SGD:EDI; MIPS:YBI047c
A:Cross-references: SGD:S0000143
A:Map position: 2L
C:Superfamily: Yeast probable calcium-binding protein YBI047c; calmodulin repeat homolog
F:167-199/Domain: calmodulin repeat homology <EF1>
F:560-576/Domain: transmembrane #status predicted <TMM>

Query Match
Best Local Similarity 6.4%; Score 403; DB 1; Length 1381;
Matches 286; Conservative 199; Mismatches 486; Indels 458; Gaps 57;

QY 3 QEPYFP-GGS-----LDVAIIVEBPAKHDOQLSLKPIAGITGQARNFFQSGLP 54
Db 109 QNPAPMOGSGATGNTNNDIPALSSNDIAKFSQLPDRAKGAQYAGDKAKDIPLKARLP 168
QY 55 QPVLQIYALADNMNDGMDQVEFSIAKMLKTLKXQGYQLPSTLPVVKQCPVALISSAPA 114
Db 169 NOTLGEIYMLCDRDASGVLDSEFTMAVYLLQLCSHPSKMTTPAVLPTQ-----219
QY 115 FCGIGIASMPPLTAVAPVPMGSIPIVVGMSPLVSSVPPAAVPLANGAPPVIOPLPAPAH 174
Db 220 -----LMDSI-----RLPEVVVNGPNRRTPLSAST 245
QY 175 PAATWPKSSFSRSGPSQOLTKLOKQSFVVASAPPAEWAPOSSRLRYOLFNSHDK 234
Db 246 GVSSILTRHSTISRLSTGAFSNA-----ASDWSLSEPKKQOPDAIPDSIDK 290
QY 235 TMSGHLPQARTIMQSSLPQAOIASIWNLSIDDDGKLTAEFFLAILIVAMSGOPLPVLP 294

```

```

Db      291 QHAGSLSSAVLVPFLLSRNLQETLATITMDLADHNNMFKLEBALMFLIOKNAIVE 350
Qy      295 LPPVLPE-----YIPSFRRVRSGSGSVISSS-----324
Db      351 LPDVIPENLLOSALGLYPPNPLPQOQSAPQAIIPSRASKPSLDPMQVAPAVNTQPT 410
Qy      325 VDQRLPE-----EPG-----SEDEQP 341
Db      411 VPQVLPQSNNGSLNDLLALNPSFSPPTAQVYVQNTNNSFYDNNNGQATLQOQOP 470
Qy      342 EKLVPTEDEKKRENFERSVELEKRRQALLQCKEKEERL-----382
Db      471 QQPPLTLHSSGGLKFTPTS-----NFGSIIKEPEBEQRLRESSTFSAQPPVPKIAS 526
Qy      383 -----382
Db      527 SPVKRTASTLLPQVNPFSVPMAGAAATGAAGAAGAAAGASAFRSNNNAEK 586
Qy      383 -----AQLEBAQEKERERQOEBAKQOLELEKQERLEQREBERKEIER 432
Db      587 QDLFADGASAKLSNATTEMANLSNQVNSLSKQASITNDKSRATQELKRVTEMKNSIQI 646
Qy      433 REAAKRELEEROLEMEHNRROELNQNKKEQETVVLKARKTLEFELEALNDKQOLE 492
Db      647 KLNMRSTHDQNVKQTEQLEAVL--QVKNENE-----TLAQGLAVSEANYHAAE 694
Qy      493 GKLODIRCLATQROEISTNKSRLAETI---THLQOQLOESQOMGLRIPEKQILS 548
Db      695 SKLNE---LTTDQESQTKNAELKEQITNLSMTASIQSQINERQO-----737
Qy      549 DOLKQVQNSLHRDILTLKRALEAKELARQOLREQLREBERFSTKQLEIDVENQOKE 608
Db      738 -QVKNQ-----ERSMVDVNSKQLELNQVYVANLQKEIDGLG-----EKISVYLTQKE 783
Qy      609 LREIHSKQOLQKORSLEAARLKQKEQ--RKSLEY---EKQKEDAQRRQERQKQMLBHV 663
Db      784 LND-----YKQTVBEQHAQQAQKQODLSNKQOTDLTDREKQLEBNRQIDEBENLYHQHV 837
Qy      664 QOEQOPRRPKRPHEDRLKREDSVKKEABERAK-----PEWQKQSLFHPQEP-- 713
Db      838 SKLQK-----MFDLSQKAKSFEKADQELKERNIYANNVRELSERQMLN-AMQOLPED 890
Qy      714 -----AKIATQAPMT---TEKGPL---TIS-----AESVAVVYVRYRLALYPESSHPI 757
Db      891 AKDITAKSASNTDTTKKATSRKGNVHEDTVSKFVETTVENSNNLVNRYKDEKERTES 950
Qy      758 TIQPGDI--VMVDESQTEGPGWLGELKGTGWFNPANYAEKIPE-----NEVPTP-AKPV 809
Db      951 DVFPDQVPTLIGSQSSENNANTNGQSGNETANP-NLNETLSDRDQDLNEVGIPIRQSL 1009
Qy      810 TDLTSAPPKALARETPAPRLPYTSEPTTPNNMADFSSSTWSSNEKEPETDNQWTMAQ 869
Db      1010 TSSVANNAPO--SVRD-DVELPETLEERDTI--NNTANRONT-----GNLSHIGEWENTPAT 1062
Qy      870 PSLTPSAGQLRQBAFAPATATGSSPSPVLG--QGEVYEGLOAQ--ALYPMWAKKXN- 923
Db      1063 ASTDVL-----NETTEVIEDGSTITKANSNEDESVSIOESPKISAOP-KAKTINE 1114
Qy      924 -----HLNFKSDVITVLEQDMMWFGEVQOQKQMPKSYVKLISGP-----965
Db      1115 EFPPIQELHIDESSSDDDE---FEDTRE---ISATYKTLQTYVNAQPTSSLIH 1166
Qy      966 ---VAKSTSIDTGPESPASLKRVASPAKAPAI-P--GEEFTAMTYESSEQGLTFQOQD 1020
Db      1167 TEQVIKYPAFGTSPSHNEGNSKKAJSTNSILPVKDEFDDEFAGL-----EQAAVEEDNG- 1219
Qy      1021 VIVVTKQGDWWTGTVGDKSGVFPSPNYVRLKDSF-----GSGTACKGSLQKREI 1071
Db      1220 -----ADSESEFENVANAGSBEQETIDHKLDELQNNATGTLTSSNPTI--PKQV 1272
Qy      1072 AQVIVATATGPEQGLTLAGQULIRKKNPGGMMGELOARQKQKQKQIFNPANYKLLSP 1131

```

```

Db      1273 QQ-----QSTSDPAQV-----SNDEWD-EIFA-----GF-----1295
Qy      1132 GTSKITPTEL--PKTAVPANCQVIGMVDYTAQDD--ELAFSGQILN 1176
Db      1296 GNSKAEPTKATPSPSQOP-----IPLKNDPIVDASLSKGPLVN 1334

RESULT 10
S28589
trichohyalin - rabbit
C:Species: Oryctolagus cuniculus (domestic rabbit)
C:Date: 12-Mar-1993 #sequence_revision 01-Mar-1996 #text_change 22-Jun-1999
C:Accession: S28589
R:Fieltz, M.J.; Rogers, G.E.
submitted to the EMBL Data Library, December 1992
A:Description: Examination of the gene encoding rabbit trichohyalin.
A:Reference number: S28589
A:Accession: S28589
A:Molecule type: DNA
A:Residues: 1-1407 <FIE>
A:Cross-references: EMBL:Z19092; NID:G1746; PID:CAA79519.1; PID:G1747
C:Comment: Trichohyalin is a protein of the medulla of the hair and of the inner root she
Covalent modifications to this protein include conversion of arginine to citrulline and t
A:introns: 46/3
C:Superfamily: trichohyalin; calmodulin repeat homology
C:Keywords: calcium binding; citrulline; EF hand; hair; tandem repeat
F:49-81/Domain: calmodulin repeat homology <EF2>

Query Match      6.24; Score 391.5; DB 1; Length 1407;
Best Local Similarity 29.78; Pred. No. 1.2e-08;
Matches 132; Conservative 97; Mismatches 140; Indels 75; Gaps 18;

Qy      326 DQRLPEBSSDEQPEKLPVTPEDKKRENFERSVELEKRRQALLQO-----375
Db      198 EEFTEBQLRRRQOELKRELEBEOQRERRERHERALQEBEQLRQRWRBEPREOQ 257
Qy      376 -----RKEORLQLEBAQERKERERQOEBAKQOLELEKQERLEBEREBER 426
Db      258 QLRLEBIEIRREOQLREBERREOQLRREQLREBERREOQLRELEBIEIRREBERGLEOE 317
Qy      427 RKE-----IERRE-AAKREL---ERQOLEMEHNRROELNQNKQO---EGTVLTKAR 473
Db      318 RREQLREBERREOQLRELEBIEIRERQRLQOE-ERBQGLAEVREBARRGSLTRRW 376
Qy      474 RKTLEFELEALNDKQKQLEGLQDIRCLATQROEISTNKSRLAETIHLQOQLO-- 531
Db      377 QROLESEKGAQSKVYSRPRROEQLRQDERQO---RQERRELEEQARRQOQOQAE 432
Qy      532 -ESQOMGLR-----IPEKQILSDOLKQVOO-----NSLHDSLLTLKRALEAKELARQ 580
Db      433 EESERRORLSARSLERQRLAEERQOEORFEEBEOQRERKQEOLOFLEBEOQLRRE 492
Qy      581 LREQLDE---VERETRSLQEI---DVFNQOKE--LREIHS-----KQOLQKORS 623
Db      493 RAQQLQEDSFQDERRRRRQOEORPGQTMWQOQEBQRRRHRLYAKPGQOQLREBE 552
Qy      624 LEAARLKQKEQERKSLLEK--QKEDAKRRQVEDKQK--LEHVQOEOPRRPHEDR 679
Db      553 LQREKRRQ-EREREYREBEKQREDEKRRQEREROYRELEBLRQEOULDRKLRBEQ 611
Qy      680 L---KREDSVKKEABERAKPEM 700
Db      612 LQREBERRLRQERERKLRBEQ 635

RESULT 11
A45592
liver seage antigen USA-1 - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 22-Nov-1993 #sequence_revision 02-Dec-1994 #text_change 09-Jun-2000
C:Accession: S24597; A45592; S29397; S34842; B45592; C45592; D45592
R:Zhu, J.; Hollingdale, M.

```


Db 804 ARPLRERERQQLAEERQREQ 826

RESULT 13

A40691
C:Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)
C:Date: 21-Sep-1993 #sequence_revision 01-Mar-1996 #text_change 22-Jun-1999
C:Accession: A40691; A34209; S32633
R:Field: M.J.; McLaughlin, C.J.; Campbell, M.T.; Rogers, G.E.
J:Cell Biol. 121, 855-865, 1993
A:Title: Analysis of the sheep trichohyalin gene: potential structural and calcium-binding
A:Reference number: A40691; MUID:93260018; PMID:7684041
A:Accession: A40691
A:Molecule type: DNA
A:Residues: 1-1549 <F1E>
A:Cross-references: EMBL:Z18361; NID:9295940; PIDN:CA79165.1; PID:9295941
C:Comment: Trichohyalin is a protein of the medulla of the hair and of the inner root sh
Covariant modifications to this protein include conversion of arginine to citrulline and
C:Genetics:
A:introns: 46/3
A:Note: single copy gene
C:Superfamily: trichohyalin; calmodulin repeat homology
C:Keywords: calcium binding; citrulline; EF hand; hair; tandem repeat
F:49-81/Domain: calmodulin repeat homology <EF2>
F:387-851/Region: 28-residue repeats
F:886-1519/Region: 23-residue repeats

Query Match 6.0%; Score 375.5; DB 1; Length 1549;
Best Local Similarity 27.6%; Pred. No. 5.9e-08;
Matches 116; Conservative 101; Mismatches 114; Indels 89; Gaps 16;
332 EPSSSEDOPEKKLPVTFE-----DKRENFERGSVELEKRRQALLQOQKEORLAOR 387
Db 383 KPAOREQVREBQRLKKEKIQREKRRQREKQVREVLQNEF--ERIQREBQLOREER 440
Qy 388 AEORKEERQ-----EQAKROLEKQ-----LEKORELEROREER 426
Db 441 EKRRERERQYLEKVELMEBQLOREBERERQREKQYLEKVELREBQLOREERKR 500
Qy 427 RKEIERREAAKRELERQOLEMERNRQELLNQRNKEBEGTVLKARKTLEFELALND 486
Db 501 ROEBERQYLEKVELQREERERK-----RROERE-----ROYLE----- 538
Qy 487 KQHOLEGLQDRIKRLATQROEISTNKSRELIAETIHLQ--QOLOESQMLRLIPEK 544
Db 539 -KVLQEEQ-----LQREKERERKQYLEKVELQREBQLOREKQREK 590
Qy 545 QILSDQLKQVQNSLHRDLSLLTKRALAKELARQOLEQULDEVERTSKLOEDIVFNN 604
Db 591 QYL-EKVELQEEBQLOREERERKQREKQREKQYLEKVELQEE--EYVQOREK-----RR 641
Qy 605 QLKRLREHSKQQLQKQSLAEALKQEKQKSLLEK-----QKEDAQ 650
Db 642 QERROYLEKELQREBQLOREBQLOREERERKQREKQREKQYLEKVELQREBQLOREER 701
Qy 651 RVQERDKQML--EYVQOREKPRPKPHEDLKRKEDSVRKKAER--AKPMQDKSRL 706
Db 702 RROERROYLEKELQREBQLOREBQLOREERERKQREKQREKQYLEKVELQREBQLOREER 758

RESULT 14

T15597

hypothetical protein C25A11.4b - Caenorhabditis elegans

C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: T15597
R:Favell, T.
submitted to the EMBL Data Library, October 1995
A:Description: The sequence of C. elegans cosmid C25A11.
A:Reference number: Z18375
A:Accession: T15597
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-737 <F8V>
A:Cross-references: EMBL:U39650; NID:91049376; PID:91049381; PIDN:AAA80392.1; CESP:C25A11
A:Experimental source: strain Bristol N2
C:Genetics:
A:Gene: CESP:C25A11.4b
A:introns: 65/3; 88/3; 126/3; 433/2; 494/2; 711/2
Query Match 5.7%; Score 359.5; DB 2; Length 737;
Best Local Similarity 23.6%; Pred. No. 1.1e-07;
Matches 179; Conservative 142; Mismatches 270; Indels 169; Gaps 32;
Qy 181 KSSGFSRSGSGQLNTKIQAGQFDVASAPPAEMAVPOSSRLKYQLFNS----HDKT 235
Db 7 QSISSRHEDLSBHATSRSTVREIPVHAPSTAP---SHSSVEEYHMMPTTSTYHHVET 63
Qy 236 MSG--HITGQATILMQS--SLPQALASINLSDLDQGLTAEEFILAMHLIDVAMS 291
Db 64 PSEBYRREVTMTITLRSSTALSQTP----- 91
Qy 292 GQPLPVLPEYIP-PSFRVRSGSGSVTSSSVQDL-----PEPSSEDE---Q 339
Db 92 GRPASPL--DRYLPYPTTTTSGDGT-RBEKTVDYKYVHNDIEQEKRIEDQARRQ 148
Qy 340 QPEKKLPVTEFDKRENFERGSVELEKRRQ-----ALLQQRKQERLAQERAEQR 392
Db 149 QEBQDRREDNRRIQAQREHQMELRQQLSERALLERERADERIQQ--ERLIRQ 207
Qy 393 KEERERQ-----EAKROLEKQOLEKRELERQREERERKEIEREAAKELER 442
Db 208 REKKRREREMWRLSIRLAEEAEALARRALEKER-IDREAAEERKTEWERERARLER 266
Qy 443 QROLMEERNRQ-----ELNQNKKEQGTV---VLKARKTLEFELALNDKKHOLEGL 495
Db 267 ER-LIEERROKKAETIRIRERERHRIETIRIKRIRIRERERREKKAED--- 322
Qy 496 QDIRCLATQROEISTNKS-REIRIAETIHLQOQLOESQMLRLIPEKQILSDQLKQV 554
Db 323 -----RLRERLEIERIERERRELEAREQELIQREAEER-----ERQRLDEARER 371
Qy 555 -QQNSLHRDLSLLTKRALAKELARQOLEQULDEVERTSKLOEDIVFNNQ--KEIREI 612
Db 372 RREERERREAEADVADVHQREERERLKRQREAEERLEIRIQOQIDMERIDAERER 431
Qy 613 HSKQOLQKSL--EAARKQEKQKSL--ELEKQEDQORQVQERDKQMLSHVQEEQ 668
Db 432 ERKEERERRELEIRIAARKKKAEDRDRLDEMERERREERERREKREBRIRIAEKER 491
Qy 669 PRPKPHEE-----DLKREDSVRYKKAERAKPEMDQKSL 706
Db 492 KRQOEEBEEIARINELQRAAARQANMELDRQORDELDRK-AQELSEREMREKERRD 550
Qy 707 FHHQEPAKLATQAPMSTTEKGPLTISAQSVYVYRALYPPESSHDITIQPDIVM 766
Db 551 REBANEAQALADLLERHNLQJRENERRAVAVERANNRLE--DRSRDL-----DIV 603
Qy 767 VDSQTEPEPMLGGLKGTGMPFANYAEKIPNEVPTPAKPTDILTSAPAKLARETP 826
Db 604 RERSKEQ-----FELEKRLIAEKEMNRKK--NHLSSG-----TLAKLT 643
Qy 827 APPLVTSSEPTTP-----NNWAFSSSTWSSS 854
Db 644 QPMVYTTREDEVTTKVERQVIERIDRNWVWEDVPAPSQS 683

RESULT 15

T15598
hypothetical protein C25A11.4a - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999

C:Accession: T15598

R: Favello, T.
submitted to the EMBL Data Library, October 1995

A:Description: The sequence of C. elegans cosmid C25A11.

A:Reference number: Z18375

A:Accession: T15598

A:Status: preliminary;

A:Molecule type: DNA

A:Residues: 1-1017 <FAN>

A:Cross-References: EMBL:U39650, NID:g1049376, PID:g1049380, PIDN:AAA80391.1; CESP:C25A1

A:Experimental source: strain Bristol N2

C:Genetics:

A:Gene: CESP:C25A11.4a

A:Introns: 65/3; 86/3; 126/3; 433/2; 494/2; 711/2; 735/1; 792/3; 833/3; 873/2

Query Match 5.7%; Score 359.5; DB 2; Length 1017;

Best Local Similarity 23.6%; Pred. No. 1.6e-07;

Matches 179; Conservative 142; Mismatches 270; Indels 169; Gaps 32;

QY 181 KSSFSRSGSGSOLNKLQKASPDVASAPPAEMAVPQSSRLKYQLFNS-----HDKT 235

DB 7 QSISSREDLSEHATSSTVEIPVHRAPSTAP--SHSVFEYHMPPTTSTYHVT 63

QY 236 MSG--HLTGPQARTILMOG--SLPQQLASIWNLSDIDQDKLTAEFFILAMHLIDVAMS 291

DB 64 PSEVYRRREVMTITITRSTALSCPTL----- 91

QY 292 GQPLPVLPEYIP--PSFRRVSGSGSVISSSVDQRL-----PEPSSDE--Q 339

DB 92 GRPASPL--DRYLPYPTTTTISGDRT--REKTVVYKTYHRDIEQERRIREDOARQ 148

QY 340 QPEKLPVTFEDKKRNFGRSGVELEKRO-----ALLEQKQERLAQLERAQER 392

DB 149 QEEQDRREDNARILAQEHQEMERLEQONLSERALARERADKERLQO--ERLLRQ 207

QY 393 KERERQEQ-----EAKROLLEKQLEKQERLEQERREKREIRERAKKELER 442

DB 208 REKKRREEDRLERSTRLEAEALARRALEKER--IDREKAEERKTMERLERERARLER 266

QY 443 QROLEWERNRO---ELLNQRNKEQGTV---VLKRRKTLFELEALNDKQQLGKTL 495

DB 267 ER--LEERQKKEKAEETERIERERREHERIEIRIKERIEREREREREKKAED--- 322

QY 496 QDIRCLATQROEISTNKS--RELRIAEITHLQOQLQESQOMGLIPKQILSDQLKOV 554

DB 323 ----RLNERLELEIERERRELEAREQEBELQREAEADR-----ERQRLDEAREM 371

QY 555 -QONSILHRDSSLTLKRALFAKELARQOLREOLDEVERSTRKLOEIDVFNOL--KELREI 612

DB 372 RRRREERRAEVLADVHRQALERERLRKQOEEREAERLIRLBOQKIDMERIDERRER 431

QY 613 HSKQQLQKQKSL--EAARLQKQEKRSKSL--ELEKQEDQORRVOERDKOMLEHVQOEEO 668

DB 432 ERKEERERREFELIEAARRKKEARDRLDEMEREVRKEERERREKRRQERRIAEKER 491

QY 669 PRPKRHEE-----DLKREDSVRKKEAEERAKPEMOKQKSL 706

DB 492 KRQEEEEEIARLNLQRAAAARQAQNAELDRQRODELDRK--AQELSEREMREKERD 550

QY 707 FHHPDEPAKLATQAPWSTKEGSLTISAQESVYVYVYALYPFESRSHDEITQPDIVM 766

DB 551 RERANDEAQLADLLERENQLIREKEREAVERANRRLE--DRRSRDL-----DHTV 603

QY 767 VDESQTGEPCWLGGELKGTGWFPANVAKIPENEVPTPAKPYTDLTSAAPAKLARETP 826

DB 604 RERSEKEQ-----FELEKRLLAEKEMNRKK--NHLISSE---TLAKLT 643

QY 827 APLVTSSEPTTP-----NNWADSSSTWPS 854

DB 644 QPMYTTREPEVTTKVERQVIERIDRWVWVEDVPYAPSQS 683

Search completed: December 4, 2003, 15:20:53
Job time : 31 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 4, 2003, 15:04:48 / Search time 15 Seconds

(without alignments)
3802.893 Million cell updates/sec

Title: US-09-674-237A-3

Perfect score: 6269

Sequence: 1 MAQFPFGSGIDVMAITVE.....VGLFSPSNVYKLTITMDPSQ 1213

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6269	100.0	1714	1 ITN1_MOUSE	Q9Z0R4 mus musculus
2	6021	96.0	1217	1 ITN1_RAT	Q9W9E9 rattus norv
3	5842.5	93.2	1721	1 ITN1_HUMAN	Q15811 homo sapien
4	5063.5	80.8	1270	1 ITN1_XENLA	Q42287 xenopus lae
5	3068	48.9	1596	1 ITN2_HUMAN	Q9R2M3 homo sapien
6	3049.5	48.6	1658	1 ITN2_MOUSE	Q9Z0U6 mus musculus
7	486.5	7.8	896	1 EP15_HUMAN	P42566 homo sapien
8	486	7.8	897	1 EP15_MOUSE	P42567 mus musculus
9	410.5	6.5	1480	1 PANT_YEAST	P32521 mus musculus
10	403	6.4	1381	1 YBET_YEAST	P34216 saccharomyc
11	391.5	6.2	1407	1 TRHY_RABIT	P37709 oryctolagus
12	381	6.1	1898	1 TRHY_HUMAN	Q07283 homo sapien
13	375.5	6.0	1549	1 TRHY_SHEEP	P22793 ovis aries
14	330	5.3	793	1 CALD_HUMAN	Q05682 homo sapien
15	330	5.3	1239	1 M4K4_HUMAN	Q95819 homo sapien
16	327	5.2	1794	1 YDC9_SCHPO	Q10172 echizosacch
17	323.5	5.2	709	1 SHP1_RAT	Q925G9 rattus norv
18	322.5	5.1	709	1 SHP1_MOUSE	Q81550 mus musculus
19	321	5.1	665	1 SHP1_HUMAN	Q96937 homo sapien
20	314.5	5.0	887	1 YLX8_CAEEL	P46504 caenorhabdi
21	309.5	4.9	2611	1 BP1E_MOUSE	Q91408 mus musculus
22	309	4.9	637	1 C2AP_MOUSE	Q91407 mus musculus
23	306.5	4.9	771	1 CALD_CHICK	P12957 gallus gall
24	304.5	4.9	585	1 INVO_HUMAN	P07476 homo sapien
25	304.5	4.9	958	1 IF3A_TOBAC	Q40554 nicotiana t
26	302	4.8	4684	1 PLB1_HUMAN	Q15149 homo sapien
27	299	4.8	1233	1 M4K4_MOUSE	P97820 mus musculus
28	298	4.8	4473	1 PLB1_CRIGR	Q91455 cricetus
29	296	4.7	4687	1 PLB1_RAT	P30427 rattus norv
30	294.5	4.7	2442	1 CEPF_HUMAN	Q9B7K3 homo sapien
31	287	4.6	639	1 C2AP_HUMAN	Q9YK66 homo sapien
32	283	4.5	733	1 VINE_MOUSE	Q91428 mus musculus
33	283	4.5	1805	1 NEST_RAT	P21263 rattus norv

34	281.5	4.5	544	1 INVO_AOTTR	P24708 actus trivi
35	280	4.5	522	1 INVO_HYLA	P17941 hylobates 1
36	280	4.5	1338	1 MYS_ABOIR	P24733 aegialacten
37	279	4.5	1023	1 GLT_DROME	P33438 drosophila
38	279	4.5	2779	1 LVA_DROME	Q86851 drosophila
39	278.5	4.4	678	1 GARP_PLAIF	P13816 plasmodium
40	278	4.4	1360	1 CING_XENLA	Q9P4D7 xenopus lae
41	277	4.4	377	1 NCK1_HUMAN	P16333 homo sapien
42	277	4.4	1208	1 DC11_DROME	P18169 drosophila
43	276.5	4.4	1197	1 CING_HUMAN	Q9P2M7 homo sapien
44	276.5	4.4	1300	1 DYNA_NEUCR	Q01397 neurospora
45	276	4.4	407	1 M21_STRPY	P50468 streptococc

ALIGNMENTS

RESULT 1	ID	ITN1_MOUSE	STANDARD	PRT	1714 AA.
AC	Q9Z0R4	Q9R143			
DT	28-FEB-2003	(Rel. 41, Created)			
DT	28-FEB-2003	(Rel. 41, Last sequence update)			
DT	15-SEP-2003	(Rel. 42, Last annotation update)			
DE	Intersect 1 (EH and SH3 domains protein 1).				
GN	ITSN1 OR ITSN OR ESEL.				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.				
OX	NCBI_Taxid=10090;				
RN	[1]				
RP	SEQUENCE FROM N.A. (ISOFORM 1).				
RX	MEDLINE=99164083; PubMed=10064583;				
RA	Sengar A.S., Wang W., Bishay J., Cohen S., Egan S.E.;				
RT	"The EH and SH3 domain ESE proteins regulate endocytosis by linking to				
RT	dynamitin and Eps15."				
RL	EMBO J. 18:1159-1171(1999).				
RL	[2]				
RP	SEQUENCE OF 966-1714 FROM N.A. (ISOFORM 2), AND SEQUENCE OF 545-599				
RP	FROM N.A. (ISOFORMS 1 AND 2).				
RC	STRAIN=129/Ola; TISSUE=Spleen;				
RA	Skripkina I.Y., Tayba L.O., Anoprienko O.V., Slavov D., Tassone F.,				
RA	Ryndtich A.V., Gardiner K.;				
RT	"Mouse homologues of human chromosome 21 genes."				
RL	Submitted (MAR-2001) to the EMBL/Genbank/DBJ databases.				
CC	- FUNCTION: Adapter protein that may provide indirect link between				
CC	the endocytic membrane traffic and the actin assembly machinery.				
CC	May regulate the formation of clathrin-coated vesicles.				
CC	- SUBUNIT: Interacts with dynamitin, SNAP-25 and SNAP-23. Clusters				
CC	several dynamitin in a manner that is regulated by alternative				
CC	splicing. Also binds clathrin-associated proteins and other				
CC	components of the endocytic machinery, such as N-WASP, Eps15 and				
CC	Stonin 2 (By similarity).				
CC	- SUBCELLULAR LOCATION: Cytoplasmic; membrane-associated protein.				
CC	- ENRICHED IN synaptoosomes (By similarity).				
CC	- ALTERNATIVE PRODUCTS:				
CC	Event=Alternative splicing; Named isoforms=2;				
CC	Comment=Additional isoforms seem to exist;				
CC	Name=1; Synonyms=Essell;				
CC	Isoid=Q9Z0R4-1; Sequence=Displayed;				
CC	Name=2;				
CC	Isoid=Q9Z0R4-2; Sequence=VSP 004296;				
CC	- TISSUE SPECIFICITY: Widely expressed. Expressed at high levels in				
CC	brain, heart and skeletal muscle.				
CC	- DOMAIN: SH3-3, SH3-4 and SH3-5, but not SH3-1 and SH3-2 domains,				
CC	bind to dynamitin (By similarity).				
CC	- DOMAIN: The KLERQ domain binds to SNAP-25 and SNAP-23 (By				
CC	similarity).				
CC	- MISCELLANEOUS: Overexpression results in the inhibition of the				
CC	transferrin uptake and the blockage of the clathrin-mediated				
CC	endocytosis.				
CC	- SIMILARITY: Contains 1 C2 domain.				
CC	- SIMILARITY: Contains 1 DBL-homology (DH) domain.				

CC -1- SIMILARITY: Contains 2 EF-hand calcium-binding domains.
 CC -1- SIMILARITY: Contains 1 PH domain.
 CC -1- SIMILARITY: Contains 5 SH3 domains.
 CC -1- SIMILARITY: Contains 2 EH domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb.ch).
 CC -----
 DR EMBL, AF132481, AAD19749.1, -
 DR EMBL, AF132478, AAD19746.1, -
 DR EMBL, AF169621, AAD48846.1, -
 DR EMBL, AF356517, AAK40228.1, -
 DR HSSP, P29354, IGFC.
 DR MGD, MGI:1338069, Ican.
 DR InterPro, IPR000008, C2.
 DR InterPro, IPR002048, EF-hand.
 DR InterPro, IPR00261, EPS15_homology.
 DR InterPro, IPR001849, PH.
 DR InterPro, IPR00219, RhoGEF.
 DR InterPro, IPR001452, SH3.
 DR Pfam, PF00168, C2_1.
 DR Pfam, PF00036, ehand, 3.
 DR Pfam, PF00169, PH, 1.
 DR Pfam, PF00621, RhoGEF, 1.
 DR Pfam, PF00018, SH3, 5.
 DR PRINTS, PR00452, SH3DOMAIN.
 DR ProDom, PD000066, SH3, 5.
 DR SMART, SM00238, C2, 1.
 DR SMART, SM00054, EH, 2.
 DR SMART, SM00027, EH, 2.
 DR SMART, SM00233, PH, 1.
 DR SMART, SM00325, RhoGEF, 1.
 DR SMART, SM00326, SH3, 5.
 DR PROSITE, PS50004, C2_DOMAIN_2, 1.
 DR PROSITE, PS50010, DH_2, 1.
 DR PROSITE, PS00018, EF_HAND, 2.
 DR PROSITE, PS50031, EH, 2.
 DR PROSITE, PS50003, PH_DOMAIN, 1.
 DR PROSITE, PS50002, SH3, 5.
 KM Endocytosis, SH3 domain; Repeat; Coiled coil; Calcium-binding;
 KM Alternative splicing.
 FT DOMAIN 21 109 EH 1.
 FT CA BIND 66 78 EF-HAND 1 (POTENTIAL).
 FT DOMAIN 221 310 EH 2.
 FT CA BIND 267 279 EF-HAND 2 (POTENTIAL).
 FT DOMAIN 326 702 LYS/LEU/GLU/ARG/GLN-RICH (KLRQ).
 FT DOMAIN 352 662 COILED COIL (POTENTIAL).
 FT DOMAIN 738 799 SH3 1.
 FT DOMAIN 906 964 SH3 2.
 FT DOMAIN 995 1053 SH3 3.
 FT DOMAIN 1067 1131 SH3 4.
 FT DOMAIN 1148 1207 SH3 5.
 FT DOMAIN 1230 1416 PH.
 FT DOMAIN 1455 1564 PH.
 FT DOMAIN 1576 1672 C2 DOMAIN.
 FT DOMAIN 321 324 POLY-SER.
 FT VANSPLC 1214 1714 Missing (in isoform 2).
 FT SEQUENCE 1714 AA; 194284 MW; 4D7AF298397860A7 CRC64;
 Query Match 100.0%; Score 6269; DB 1; Length 1714;
 Best Local Similarity 100.0%; Pred. No. 3,9e-232;
 Matches 1213; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 61 IVALADNNNDGRMDQVEFSIAMKLIKLGQYQUPSTLPYPMKQOYVAISSAPFGIGGI 120
 DB 61 IVALADNNNDGRMDQVEFSIAMKLIKLGQYQUPSTLPYPMKQOYVAISSAPFGIGGI 120
 QY 121 ASMPPLTAVAPVNGSIPIVVGMSPPVSVVPAAPVPIPLANGAPPVIOPLPAFAHPATWP 180
 DB 121 ASMPPLTAVAPVNGSIPIVVGMSPPVSVVPAAPVPIPLANGAPPVIOPLPAFAHPATWP 180
 QY 181 KSSSFSSSGPSQUNTLQKRAQSPDVASAPPAAWAPQSSRLKYROLFNSHDTMGSGL 240
 DB 181 KSSSFSSSGPSQUNTLQKRAQSPDVASAPPAAWAPQSSRLKYROLFNSHDTMGSGL 240
 QY 241 TGPAQTILMOSSLPOKQASIMNLSIDIDQKLTAEFFILAMLIDVAMSGQLPVLP 300
 DB 241 TGPAQTILMOSSLPOKQASIMNLSIDIDQKLTAEFFILAMLIDVAMSGQLPVLP 300
 QY 301 PEYIPSPFRVRSVSGMSVSISSSVQDRLPEPSSSEDEQPEKKLPYTFEDKKENFERG 360
 DB 301 PEYIPSPFRVRSVSGMSVSISSSVQDRLPEPSSSEDEQPEKKLPYTFEDKKENFERG 360
 QY 361 SVELKRRQALLLEQKREQLAQLERAQERKEREQDEAKQLELEKQLEKRELER 420
 DB 361 SVELKRRQALLLEQKREQLAQLERAQERKEREQDEAKQLELEKQLEKRELER 420
 QY 421 QREBERKEIRBRPAKRELERQRLERERROELNQRKEQEGTVVLKARRKLEFE 480
 DB 421 QREBERKEIRBRPAKRELERQRLERERROELNQRKEQEGTVVLKARRKLEFE 480
 QY 481 LEALNDKKGQLEGLQDIRCLATQROEISTNKSREIRIAEITHLQOOLQESQOMIGRL 540
 DB 481 LEALNDKKGQLEGLQDIRCLATQROEISTNKSREIRIAEITHLQOOLQESQOMIGRL 540
 QY 541 IPEKQILSDQKQVQNSLHRDLSLTLLKRALEAELARQQLREQLDEVERTRSKLEID 600
 DB 541 IPEKQILSDQKQVQNSLHRDLSLTLLKRALEAELARQQLREQLDEVERTRSKLEID 600
 QY 601 VFNNQLEKREIHSKQOLQKORSLEAARLKQKEERKSLELEKQEDAKQRRVQERDKWL 660
 DB 601 VFNNQLEKREIHSKQOLQKORSLEAARLKQKEERKSLELEKQEDAKQRRVQERDKWL 660
 QY 661 EHVQOEBQPRPKPHEEDRLKREDSVKKKEAEERAKEMODKQRLPHPHQEPKATQA 720
 DB 661 EHVQOEBQPRPKPHEEDRLKREDSVKKKEAEERAKEMODKQRLPHPHQEPKATQA 720
 QY 721 PWSITTEKGLTISAQESVKKVYRYALYPFESRSHDEITIQGDIVMDESGTGPGLG 780
 DB 721 PWSITTEKGLTISAQESVKKVYRYALYPFESRSHDEITIQGDIVMDESGTGPGLG 780
 QY 781 ELKQKTGMPFANYAEKIPENEVPTPAKPVTDLTSAAPKTLARETPAPLPYTSSEPSITP 840
 DB 781 ELKQKTGMPFANYAEKIPENEVPTPAKPVTDLTSAAPKTLARETPAPLPYTSSEPSITP 840
 QY 841 NNMADFSSTWSSSNEKRETDNMDTMAAQPSTLVPSAGQLRQSAFTPATATGSSPSPLV 900
 DB 841 NNMADFSSTWSSSNEKRETDNMDTMAAQPSTLVPSAGQLRQSAFTPATATGSSPSPLV 900
 QY 901 GQGERKVEGLQALYIPRAKKNHNLNFKSVITVLQOQDMWVGEGVQGGMPKSVK 960
 DB 901 GQGERKVEGLQALYIPRAKKNHNLNFKSVITVLQOQDMWVGEGVQGGMPKSVK 960
 QY 961 LISGPVRKSTIDTGPESPASLKRVASPAKPAIPGEFTAMTYTSSSEGGDLTFQGD 1020
 DB 961 LISGPVRKSTIDTGPESPASLKRVASPAKPAIPGEFTAMTYTSSSEGGDLTFQGD 1020
 QY 1021 VIIVTKKDDGDMWTGVDKSGVFPNSNYRLKDSGSGTAGTGSGLGKKPEIAQVIASYAA 1080
 DB 1021 VIIVTKKDDGDMWTGVDKSGVFPNSNYRLKDSGSGTAGTGSGLGKKPEIAQVIASYAA 1080
 QY 1081 TGPEQLTLAPQOLLIRKKNPGGMEGELQARKKROIGWFPANVYKLLSGTSTKITPTE 1140
 DB 1081 TGPEQLTLAPQOLLIRKKNPGGMEGELQARKKROIGWFPANVYKLLSGTSTKITPTE 1140
 QY 1141 LPKTAVQPAVCQVIGMTDYTAQNDDLAFSKGIINVLNKEDPMMWKEVSGVGLFPSN 1200

Db 1141 LEKTAIVOPAVCOVIGDYTAQNDDELAFKSGQIIIVLNKEDPDMWKGEVSGVGLFSPN 1200
 QY 1201 YVKLTTPMDPSOO 1213
 Db 1201 YVKLTTPMDPSOO 1213

RESULT 2
 ITN1 RAT
 ID ITN1 RAT STANDARD; PRT; 1217 AA.

AC Q9WVE9; Q9WVE1; STANFORD; PRT; 1217 AA.
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Intersectin 1 (EH domain and SH3 domain regulator of endocytosis 1).
 GN ITSN1 OR ITSN OR EHSN1.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
 RC TISSUE=Brain;
 RX MEDLINE=99303609; PubMed=10373452;
 RA Okamoto M., Schuch S., Suedhof T.C.;
 RT "EHSN1/intersectin, a protein that contains EH and SH3 domains and binds to dynamin and SNAP-25. A protein connection between exocytosis and endocytosis?";
 RT J. Biol. Chem. 274:18446-18454(1999).
 RL -1- FUNCTION: Adapter protein that may provide indirect link between the endocytic membrane traffic and the actin assembly machinery. May regulate the formation of clathrin-coated vesicles.
 CC -1- SUBUNIT: Interacts with dynamin, SNAP-25 and SNAP-23. Clusters several dynamin in a manner that is regulated by alternative components of the endocytic machinery, such as N-Wasp, Eps15 and Stc1n 2 (By similarity).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic; membrane-associated protein.
 CC -1- ENRICHED IN synaposomes.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Comment=Additional isoforms seem to exist;
 CC Name=1;
 CC IsoId=Q9WVE9-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q9WVE9-2; Sequence=VSP_004297;
 CC -1- TISSUE SPECIFICITY: Highly expressed in brain.
 CC -1- DOMAIN: SH3-3, SH3-4 and SH3-5, but not SH3-1 and SH3-2 domains, bind to dynamin.
 CC -1- DOMAIN: The KLERQ domain binds to SNAP-25 and SNAP-23.
 CC -1- SIMILARITY: Contains 2 EF-hand calcium-binding domains.
 CC -1- SIMILARITY: Contains 5 SH3 domains.
 CC -1- SIMILARITY: Contains 2 EH domains.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).

DR SMART; SM00027; EH; 2.
 DR SMART; SM00326; SH3; 5.
 DR PROSITE; PS00018; EF_HAND; 2.
 DR PROSITE; PS50031; EH; 2.
 DR PROSITE; PS50002; SH3; 5.
 DR Endocytosis; SH3 domain; Repeat; Coiled coil; Calcium-binding;
 KW Alternative splicing.
 FT DOMAIN 21 109 EH 1.
 FT CA BIND 66 78 EF_HAND 1 (POTENTIAL).
 FT DOMAIN 221 310 EH 2.
 FT CA BIND 267 279 EF_HAND 2 (POTENTIAL).
 FT DOMAIN 326 702 LYS/LEU/GLU/ARG/GLN-RICH (KLERQ).
 FT DOMAIN 350 670 COILED COIL (POTENTIAL).
 FT DOMAIN 737 803 SH3 1.
 FT DOMAIN 910 968 SH3 2.
 FT DOMAIN 959 1057 SH3 3.
 FT DOMAIN 1071 1135 SH3 4.
 FT DOMAIN 1152 1211 SH3 5.
 FT DOMAIN 321 324 POLY-SER.
 FT VARSPLC 1003 1073 Missing (in isoform 2).
 FT /FTID=VSP_004297.
 SQ SEQUENCE 1217 AA; 137154 MW; 6C1323BAE5A5B34B CRC64;
 Query Match 96.0%; Score 6021; DB 1; Length 1217;
 Best Local Similarity 95.7%; Pred. No. 7,9e-223;
 Matches 1166; Conservative 19; Mismatches 27; Indels 6; Gaps 2;

QY 1 MAQFTPTGGSLDVAITVEERAKHDQFLSLKPIAGFTTGDQANFFQSGLPQVLAQ 60
 Db 1 MAQFTPTGGSLDVAITVEERAKHDQFQSLKPIAGFTTGDQANFFQSGLPQVLAQ 60
 QY 61 IVALADNMNDGMDOVERSIMKIKLKLGQVLPSTLPVWKKOOPVAISSAPAGIGGI 120
 Db 61 IVALADNMNDGMDOVERSIMKIKLKLGQVLPSTLPVWKKOOPVAISSAPAGIGGI 120
 QY 121 ASMPPLTVAPVPMQSIPIVGMSPPLVSSVPPAAPPVLANGAPVPIQPLPAFAHPAATWP 180
 Db 121 ASMPPLTVAPVPMQSIPIVGMSPPLVSSVPPAAPPVLANGAPVPIQPLPAFAHPAATWP 180
 QY 121 AGMPPLTVAPVPMQSIPIVGMSPPLVSSVPPAAPPVLANGAPVPIQPLPAFAHPAATWP 180
 Db 121 AGMPPLTVAPVPMQSIPIVGMSPPLVSSVPPAAPPVLANGAPVPIQPLPAFAHPAATWP 180
 QY 181 KSSFSRSGSGSLNTKLQKQSPVAVASAPPAEWAHPQSSRLKXKRLFNSHDKTWSGHL 240
 Db 181 KSSFSRSGSGSLNTKLQKQSPVAVASAPPAEWAHPQSSRLKXKRLFNSHDKTWSGHL 240
 QY 181 KSSFSRSGSGSLNTKLQKQSPVAVASAPPAEWAHPQSSRLKXKRLFNSHDKTWSGHL 240
 Db 181 KSSFSRSGSGSLNTKLQKQSPVAVASAPPAEWAHPQSSRLKXKRLFNSHDKTWSGHL 240
 QY 241 TGPQARTILMOSLSLPQALASLWNLSDIDQDKLTAEFFILAMHLIDVAMSGQPLPVLP 300
 Db 241 TGPQARTILMOSLSLPQALASLWNLSDIDQDKLTAEFFILAMHLIDVAMSGQPLPVLP 300
 QY 301 PEYIPSPFRARVSGSGMVISVSSVVDORLPPEPSSSEDOPEKPLPVTFEDKRENPFRRG 360
 Db 301 PEYIPSPFRARVSGSGMVISVSSVVDORLPPEPSSSEDOPEKPLPVTFEDKRENPFRRG 360
 QY 301 PEYIPSPFRARVSGSGMVISVSSVVDORLPPEPSSSEDOPEKPLPVTFEDKRENPFRRG 360
 Db 301 PEYIPSPFRARVSGSGMVISVSSVVDORLPPEPSSSEDOPEKPLPVTFEDKRENPFRRG 360
 QY 361 SVELERKROALLQQRKEQERLAQLEPAEOERKEREROEAKROLLEKOLEKRELER 420
 Db 361 SVELERKROALLQQRKEQERLAQLEPAEOERKEREROEAKROLLEKOLEKRELER 420
 QY 361 NLELEKROALLQQRKEQERLAQLEPAEOERKEREROEAKROLLEKOLEKRELER 420
 Db 361 NLELEKROALLQQRKEQERLAQLEPAEOERKEREROEAKROLLEKOLEKRELER 420
 QY 421 QREERREKIERREAKELEROROLEMERNRROELNQRNKGQGVVVKARRKTLLEFE 480
 Db 421 QREERREKIERREAKELEROROLEMERNRROELNQRNKGQGVVVKARRKTLLEFE 480
 QY 421 QREERREKIERREAKELEROROLEMERNRROELNQRNKGQGVVVKARRKTLLEFE 480
 Db 421 QREERREKIERREAKELEROROLEMERNRROELNQRNKGQGVVVKARRKTLLEFE 480
 QY 481 LEALNDKGOLEGLKQDTRCRLATRORETESNRSRELRIRIETHLQOOLQESQOMLRL 540
 Db 481 LEALNDKGOLEGLKQDTRCRLATRORETESNRSRELRIRIETHLQOOLQESQOMLRL 540
 QY 481 LEALNDKGOLEGLKQDTRCRLATRORETESNRSRELRIRIETHLQOOLQESQOMLRL 540
 Db 481 LEALNDKGOLEGLKQDTRCRLATRORETESNRSRELRIRIETHLQOOLQESQOMLRL 540
 QY 541 IPEKQILSDOLKOVQNSLHRDSTLLTKRALEAKELAQOLREOLDVEVERSTRKLOEID 600
 Db 541 IPEKQILSDOLKOVQNSLHRDSTLLTKRALEAKELAQOLREOLDVEVERSTRKLOEID 600
 QY 541 IPEKQILSDOLKOVQNSLHRDSTLLTKRALEAKELAQOLREOLDVEVERSTRKLOEID 600
 Db 541 IPEKQILSDOLKOVQNSLHRDSTLLTKRALEAKELAQOLREOLDVEVERSTRKLOEID 600
 QY 601 VFNNOLEKLEIRHSQOLQKQSLLEAALKQKQERKSLLEKQKEDQKRVQEDKQWL 660
 Db 601 VFNNOLEKLEIRHSQOLQKQSLLEAALKQKQERKSLLEKQKEDQKRVQEDKQWL 660
 QY 601 VFNNOLEKLEIRHSQOLQKQSLLEAALKQKQERKSLLEKQKEDQKRVQEDKQWL 660
 Db 601 VFNNOLEKLEIRHSQOLQKQSLLEAALKQKQERKSLLEKQKEDQKRVQEDKQWL 660
 QY 661 EHVQOEQPRPKPHEEDRLKREDSVRKGEAEERKAPPMQDKOSLFPHPQEPATLQTA 720
 Db 661 EHVQOEQPRPKPHEEDRLKREDSVRKGEAEERKAPPMQDKOSLFPHPQEPATLQTA 720

Db 661 EHVQOEQORPRKPHEDDKLAKREDSVKKKAEEBAKPEVQDKSRLFFHQEPAPKA-QA 719
 Qy 721 PMSTTEKPLTISAOESVKVYVYALYFESRSHEITIQGCDIVM-----VDASQCEP 775
 Db 720 PMFTTEKPLTISAOESAKVYVYALYFESRSHEITIQGDIWMKGEVNDSSQGE 779
 Qy 776 GMLGELGKGTGWPFPANAEKIPENEVPTPAKPYDLTSAAPKLAETPAPLPTVSS 835
 Db 780 GMLGEPKGTGWPFPANAEKIPENEITPAKPYDLTSAAPKLAETPAPLPTVSS 839
 Qy 836 PSTTPNNMADPSSSWSSNEKEPTDNDVTAADPSLTVPAGQIRORSAPTPATGSS 895
 Db 840 PSTTPNNMADPSSSWSSNEKEPTDNDVTAADPSLTVPAGQIRORSAPTPATGSS 899
 Qy 896 PSPVIGQEKVEGLOAOLYPMRAKKNHLPNNSDVTVLEODMMFGEVQCKGFP 955
 Db 900 PSPVIGQEKVEGLOAOLYPMRAKKNHLPNNSDVTVLEODMMFGEVQCKGFP 959
 Qy 956 KSYVGLISGPRKSTIDTGPTEPSASLKRVAAPAKPAIPEEFIAMYTESSBOGDLT 1015
 Db 960 KSYVGLISGPRKSTIDTGPTEPSASLKRVAAPAKPAIPEEFIAMYTESSBOGDLT 1019
 Qy 1016 FOQGDVIVVTKKDGDMWTGTVGDKSGVFPSSNVYRLKDSGSGTAGKTSLGAKPEIAQVI 1075
 Db 1020 FOQGHVIVVTKKDGDMWTGTVGDSGVPSSNVYRLKDSGSGTAGKTSLEKPEIAQVI 1079
 Qy 1076 ASYATGGEOLTLAAGQILIRKKNPGGMEGELQARKKQIQGFPANVYKLSPGTSK 1135
 Db 1080 ASYATGGEOLTLAAGQILIRKKNPGGMEGELQARKKQIQGFPANVYKLSPGTSK 1139
 Qy 1136 ITPTLPPTAVOPACQVIGMYDYAQNDELAFESKQIINVLNKEDPDMWKGEVSGQVG 1195
 Db 1140 ITPTLPPTAVOPACQVIGMYDYAQNDELAFESKQIINVLNKEDPDMWKGEVSGQVG 1199
 Qy 1196 LFPSNVYKLTDTMDPSQ 1213
 Db 1200 LFPSNVYKLTDTMDPSQ 1217

RESULT 3

ID ITN1 HUMAN STANDARD; PRT; 1721 AA.
 AC Q15811; Q95216; Q9UK60; Q9UNK1; Q9UNK2; Q9UQ92;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Intersectin 1 (SH3 domain-containing protein 1A) (SH3P17).
 GN ITSN1 OR ITSN OR SH3D1A.
 OS Homo sapiens (Human).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 CX NCBI_TaxID:9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
 RC TISSUE=Fetal brain;
 RX MEDLINE=990117974; PubMed=9799604;
 RA Antonarakis S.E.;
 RA Guipponi M., Scott H.S., Chen H., Schebesta A., Rossier C.,
 RA Antonarakis S.E.;
 RT "Two isoforms of a human intersectin (ITSN) protein are produced by
 RT brain-specific alternative splicing in a stop codon."
 RL Genomics 53:369-376(1998).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
 RX MEDLINE=99415290; PubMed=10482960;
 RA Pucharcos C., Fuentes J.-J., Casas C., de la Luna S., Alcantara S.,
 RA Arbones M.L., Soriano E., Estivill X., Pritchard M.;
 RT "Alu-splice cloning of human intersectin (ITSN), a putative
 RT multivalent binding protein expressed in proliferating and
 RT differentiating neurons and overexpressed in Down syndrome."
 RL Eur. J. Hum. Genet. 7:704-712(1999).
 RN [3]
 RP SEQUENCE OF 963-1721 FROM N.A. (ISOFORM 3).
 RC TISSUE=Brain;

RA Teyba L.O., Kvaasha S.M., Skripkina I.Y., Anoprienko O.V., Slavov D.,
 RA Tassone F., Rynditch A.V., Gardiner K.;
 RT "Mouse homologs of human chromosome 21 genes";
 RL Submitted (Aug-1999) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE OF 620-1721 FROM N.A. (ISOFORM 3).
 RC TISSUE=Bone marrow;
 RX MEDLINE=98294438; PubMed=9630982;
 RA Sparks A.B., Hofman N.G., McConnell S.J., Fowlkes D.M., Kay B.K.;
 RT "Cloning of ligand targets: systematic isolation of SH3 domain-
 RT containing proteins";
 RL Nat. Biotechnol. 14:741-744(1996).
 RN [5]
 RP ALTERNATIVE SPLICING.
 RC TISSUE=Brain, and Fetal liver;
 RX MEDLINE=21548828; PubMed=11690630;
 RA Pucharcos C., Casas C., Nadal M., Estivill X., de la Luna S.;
 RT "The human intersectin genes and their spliced variants are
 RT differentially expressed";
 RL Biochim. Biophys. Acta 1521:1-11(2001).
 RN [6]
 RP GENE MAPPING.
 RX MEDLINE=98127038; PubMed=9465890;
 RA Chen H., Antonarakis S.E.;
 RT "The SH3D1A gene maps to human chromosome 21q22.1-->q22.2";
 RL Cytogenet. Cell Genet. 78:213-215(1997).
 CC -1- FUNCTION: Adapter protein that may provide indirect link between
 CC the endocytic membrane traffic and the actin assembly machinery.
 CC May regulate the formation of clathrin-coated vesicles. Isoform 1
 CC could be involved in brain-specific synaptic vesicle recycling.
 CC -1- SUBUNIT: Interacts with dynamin, SNAP-25 and SNAP-23. Clusters
 CC severl dynamin in a manner that is regulated by alternative
 CC splicing. Also binds clathrin-associated proteins and other
 CC components of the endocytic machinery, such as N-WASP, Eps15 and
 CC Secin 2 (By similarity).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic; membrane-associated protein.
 CC Enriched in synaptosomes (By similarity).
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event-Alternative splicing: Named isoforms=4;
 CC Comment=Additional isoforms seem to exist. Alternative splicing
 CC affects domains involved in protein recognition and thus may
 CC play a role in selecting specific interactions;
 CC Name=1; Synonym=Long, ITSN-1;
 CC IsoId=Q15811-1; Sequence=Displayed;
 CC Name=2; Synonym=Short, ITSN-S;
 CC IsoId=Q15811-2; Sequence=VSP_004295;
 CC Name=3; Synonym=Short 2, SH3P17;
 CC IsoId=Q15811-3; Sequence=VSP_004293, VSP_004294, VSP_004295;
 CC Name=4;
 CC IsoId=Q15811-4; Sequence=VSP_004294;
 CC -1- TISSUE SPECIFICITY: Ubiquitous in adult and fetal tissues, except
 CC isoform 1 which is expressed almost exclusively in the brain.
 CC Highly expressed in skeletal muscle, heart, spleen, ovary, testis
 CC and all fetal tissues tested. Expressed at lower levels in thymus,
 CC blood, lung, liver and pancreas. Isoform 1 is expressed in all
 CC brain regions; not expressed in the spinal cord.
 CC -1- DOMAIN: SH3-3, SH3-4 and SH3-5 but not SH3-1 and SH3-2 domains,
 CC bind to dynamin (By similarity).
 CC -1- DOMAIN: The KLRQ domain binds to SNAP-25 and SNAP-23 (By
 CC similarity).
 CC -1- DISEASE: Overexpressed in brain from Down syndrome fetuses
 CC suggesting a gene dosage-dependent contribution to the
 CC abnormalities of Down syndrome.
 CC -1- SIMILARITY: Contains 1 C2 domain.
 CC -1- SIMILARITY: Contains 1 DBP-homology (DB) domain.
 CC -1- SIMILARITY: Contains 2 EF-hand calcium-binding domains.
 CC -1- SIMILARITY: Contains 1 PH domain.
 CC -1- SIMILARITY: Contains 5 SH3 domains.
 CC -1- SIMILARITY: Contains 2 EH domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL, AF064244; AAC78611.1; -;
 DR EMBL, AF064247; AAC80437.1; -;
 DR EMBL, AF064245; AAC80437.1; JOINED.
 DR EMBL, AF064246; AAC80437.1; JOINED.
 DR EMBL, AF064243; AAC78610.1; -;
 DR EMBL, AF114488; AAD29953.1; -;
 DR EMBL, AF114487; AAD29952.1; -;
 DR EMBL, AF180522; AAD53183.1; -;
 DR EMBL, U61166; AAC50592.1; ALT_INIT.
 DR PDB, 1K11; 29-MAY-02.
 DR Genew: HGNC:6183; ITSN1.
 DR MIM, 602442; -;
 DR GO; GO:0005509; F:calcium ion binding activity; NAS.
 DR GO; GO:0005085; F:guanylyl-nucleotide exchange factor activity; NAS.
 DR InterPro; IPR000008; C2.
 DR InterPro; IPR002048; EF-hand.
 DR InterPro; IPR000261; Eps15_homology.
 DR InterPro; IPR001018; Neu_cyt_fac2_2.
 DR InterPro; IPR001849; PH.
 DR InterPro; IPR00219; RhogEF.
 DR InterPro; IPR001452; SH3.
 DR Pfam; PF00036; eHand; 3.
 DR Pfam; PF00169; PH; 1.
 DR Pfam; PF00621; RhogEF; 1.
 DR Pfam; PF00018; SH3; 5.
 DR PRINTS; PR00499; P67PHOX.
 DR PRINTS; PR00452; SH3DOMAIN.
 DR Prodom; P000066; SH3; 5.
 DR SMART; SM00239; C2; 1.
 DR SMART; SM00054; EFh; 2.
 DR SMART; SM00027; EH; 2.
 DR SMART; SM00233; PH; 1.
 DR SMART; SM00325; RhogEF; 1.
 DR SMART; SM00326; SH3; 5.
 DR PROSITE; PS50004; C2_DOMAIN_2; 1.
 DR PROSITE; PS50010; DH_2; 1.
 DR PROSITE; PS00018; EF_HAND; 2.
 DR PROSITE; PS50031; EH; 2.
 DR PROSITE; PS50003; PH_DOMAIN; 1.
 DR PROSITE; PS50002; SH3; 5.
 DR Endocytosis; SH3 domain; Repeat; Coiled coil; Calcium-binding;
 KM Alternative splicing; 3D-structure.
 FT DOMAIN 21 109
 FT CA BIND 66 77
 FT DOMAIN 221 310
 FT CA BIND 267 278
 FT CA BIND 366 702
 FT DOMAIN 351 705
 FT DOMAIN 740 806
 FT DOMAIN 913 971
 FT DOMAIN 1002 1060
 FT DOMAIN 1074 1138
 FT DOMAIN 1155 1214
 FT DOMAIN 1237 1423
 FT DOMAIN 1462 1571
 FT DOMAIN 1583 1679
 FT VARSPLIC 770 774
 FT VARSPLIC 1006 1076
 FT VARSPLIC 1221 1721
 FT CONFLICT 114 114
 FT CONFLICT 1088 1088
 FT CONFLICT 1109 1109
 FT CONFLICT 1137 1137
 FT CONFLICT 1361 1361

FT CONFLICT 1474 1474 N -> S (IN REF. 2).
 SQ SEQUENCE 1721 AA; 195532 MW; 8189DDDA0F145B4B5 CRC64;
 Query Match 93.2%; Score 5842.5; DB 1; Length 1721;
 Best Local Similarity 92.5%; Pred. No. 7.1e-216;
 Matches 1128; Conservative 39; Mismatches 46; Indels 7; Gaps 3;

QY 1 MAQPTPFGGSLDVVAITVEERAKHDOQLSKPIAGFTGQANFFQSLPOPVLAQ 60
 DB 1 MAQPTPFGGSLDVVAITVEERAKHDOQLSKPIAGFTGQANFFQSLPOPVLAQ 60
 QY 61 IWLADNMNDGMDQVERSIAMKLIKLIKQYQLPSTLPVVKQOPVAISSAPAGIGI 120
 DB 61 IWLADNMNDGMDQVERSIAMKLIKLIKQYQLPSTLPVVKQOPVAISSAPAGIGI 120
 QY 121 ASMPPLTAVAPVPMGSIYVGMSPVLVSSVPPAAVPLANGAPVYIQPLPAFHPAATLP 180
 DB 121 ASMPPLTAVAPVPMGSIYVGMSPVLVSSVPPAAVPLANGAPVYIQPLPAFHPAATLP 180
 QY 181 KSSFSRSGSGSLQNTKLOKQSPVAPAPAAEAVQSSRLKXROLFNSHDKTMSGTL 240
 DB 181 KSSFSRSGSGSLQNTKLOKQSPVAPAPAAEAVQSSRLKXROLFNSHDKTMSGTL 240
 QY 241 TGPQARTILMOSLTPQAOIASIWNLSIDIDQDKLTAEETILAMHLIDVAMSGQPLPVLP 300
 DB 241 TGPQARTILMOSLTPQAOIASIWNLSIDIDQDKLTAEETILAMHLIDVAMSGQPLPVLP 300
 QY 301 PEYIPSPRRVAVSSGGMVYSSSVDOQLPEPSSSEDEQOP-EKLLPTFEDKKBENFER 359
 DB 301 PEYIPSPRRVAVSSGGMVYSSSVDOQLPEPSSSEDEQOP-EKLLPTFEDKKBENFER 359
 QY 360 GSVELKKRQALLBOORKEOERLAQERAEORKEEREOEAKROLEKORELE 419
 DB 360 GSVELKKRQALLBOORKEOERLAQERAEORKEEREOEAKROLEKORELE 419
 QY 420 RQREERKEIERREARERLEERQLEMERNRROELNORKEOEGTVLAKARKTLEF 479
 DB 420 RQREERKEIERREARERLEERQLEMERNRROELNORKEOEGTVLAKARKTLEF 479
 QY 480 ELEALNDKXKQLEGLDIDICRLATOROIESTNSKRELRIAETHTLOOQLOESQOMIGR 539
 DB 480 ELEALNDKXKQLEGLDIDICRLATOROIESTNSKRELRIAETHTLOOQLOESQOMIGR 539
 QY 540 LIPEKQIISDQKQVQNSIHRSLLTKRLAEELARQOLRELDVYERTSKLQEI 599
 DB 540 LIPEKQIISDQKQVQNSIHRSLLTKRLAEELARQOLRELDVYERTSKLQEI 599
 QY 600 DVFNNQLEKEIHHSKQOLQKORSLAEARLKQKEQERKSLLEKQEDAKORVQRRDKQW 659
 DB 600 DVFNNQLEKEIHHSKQOLQKORSLAEARLKQKEQERKSLLEKQEDAKORVQRRDKQW 659
 QY 660 LEHVQOE-EOPRPKREHEDRIKREDSYRKKEAEERAKPENQDKOSRLPHHQEPAKLAT 718
 DB 660 LEHVQOE-EOPRPKREHEDRIKREDSYRKKEAEERAKPENQDKOSRLPHHQEPAKLAT 718
 QY 719 QAPWSTERKGLTISAQSVVYVYRALYPRESRHDEITTOPIQIWM-----VESQSG 773
 DB 719 QAPWSTERKGLTISAQSVVYVYRALYPRESRHDEITTOPIQIWM-----VESQSG 773
 QY 721 QAPWSTERKGLTISAQSVVYVYRALYPRESRHDEITTOPIQIWM-----VESQSG 780
 DB 721 QAPWSTERKGLTISAQSVVYVYRALYPRESRHDEITTOPIQIWM-----VESQSG 780
 QY 774 EFGWLGELGKGTGFWPANYAEKIPENEPVTPAKVNTLTSAPAKLARETAPLAPYTS 833
 DB 774 EFGWLGELGKGTGFWPANYAEKIPENEPVTPAKVNTLTSAPAKLARETAPLAPYTS 833
 QY 781 EFGWLGELGKGTGFWPANYAEKIPENEPVTPAKVNTLTSAPAKLARETAPLAPYTS 840
 DB 781 EFGWLGELGKGTGFWPANYAEKIPENEPVTPAKVNTLTSAPAKLARETAPLAPYTS 840
 QY 834 SEPSTTPNNWADFSTWPSSSNEKETNDMTMAAQPSTLYPSAQLORSASFATATG 893
 DB 834 SEPSTTPNNWADFSTWPSSSNEKETNDMTMAAQPSTLYPSAQLORSASFATATG 893
 QY 841 SEPSTTPNNWADFSTWPSSTTWEKETNDMTMAAQPSTLYPSAQLORSASFATATG 900
 DB 841 SEPSTTPNNWADFSTWPSSTTWEKETNDMTMAAQPSTLYPSAQLORSASFATATG 900
 QY 894 SSPSPVLQGEKVBGLQALYPMWAKKDNHLPKPSVITVLEQDDMMWREVGQGM 953
 DB 894 SSPSPVLQGEKVBGLQALYPMWAKKDNHLPKPSVITVLEQDDMMWREVGQGM 953
 QY 901 SSPSPVLQGEKVBGLQALYPMWAKKDNHLPKPSVITVLEQDDMMWREVGQGM 960
 DB 901 SSPSPVLQGEKVBGLQALYPMWAKKDNHLPKPSVITVLEQDDMMWREVGQGM 960
 QY 954 FPKSYVKLISGVRKSTSIDTGPESPASLKRVASPAKPAIPGEEFTAMTYESSEOGD 1013
 DB 954 FPKSYVKLISGVRKSTSIDTGPESPASLKRVASPAKPAIPGEEFTAMTYESSEOGD 1013

```

Db 961 PFKSVYKLSGPIRKSTMSDSSSPASLKRVASPAKPVVSGEEFIAMTYESSECD 1020
Qy 1014 LTFQGGDYVITVKDGMWGTGDKSVFPSPNYRLKDSGSGTAGTSLGKKPEIAQ 1073
Db 1021 LTFQGGDYVITVKDGMWGTGDKSVFPSPNYRLKDSGSGTAGTSLGKKPEIAQ 1080
Qy 1074 VIASAAATGPPOLTLAPGQLILIRKNGPGWMEGLQARGKKROIQWFPANVVKLSPGT 1133
Db 1081 VIASATATGPPOLTLAPGQLILIRKNGPGWMEGLQARGKKROIQWFPANVVKLSPGT 1140
Qy 1134 SKITPTLPKTAQVAPVAVCVIYGYDYTAQNDDELAFSGKQIIVLNKEDPDMKGEVSGQ 1193
Db 1141 SKITPTLPKSTALAAVCQVIGYDYTAQNDDELAFNKGQIIVLNKEDPDMKGEVVGQ 1200
Qy 1194 VGLFPSPNYVKLTTPMDPSQ 1213
Db 1201 VGLFPSPNYVKLTTPMDPSQ 1220

RESULT 4
ITN1_XENLA STANDARD; PRT; 1270 AA.
AC 04287;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Intersectin 1.
GN ITSN1.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
OC Xenopodinae; Xenopus.
CX NCBI_TaxID=8335;
[1]
RP SEQUENCE FROM N.A.
RC TISSUE=Oocyte;
RX MEDLINE=99030416; PubMed=9813051;
RA Yamauchi M., Hoffman N.G., Hardison N.L., McPherson P.S.,
RA Castagnoli L., Cesareni G., Kay B.K.,
RT "Intersectin, a novel adaptor protein with two eps15 homology and five
RT src homology 3 domains."
RL J. Biol. Chem. 273:31401-31407(1998).
CC -1- FUNCTION: Adaptor protein that may provide indirect link between
CC the endocytic membrane traffic and the actin assembly machinery.
CC May regulate the formation of clathrin-coated vesicles.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic; membrane-associated protein.
CC -1- SIMILARITY: Contains 2 EF-hand calcium-binding domains.
CC -1- SIMILARITY: Contains 5 SH3 domains.
CC -1- SIMILARITY: Contains 2 EH domains.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL, AF032118; AAC73068.1; -
CC PIR, T09194; T09194.
CC HSSP, P29355; 1SEM.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR000261; Eps15_homology.
DR InterPro; IPR001018; Neu_cyt_fact_2.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00036; ehand; 2.
DR Pfam; PF00018; SH3; 5.
DR PRINTS; PR00499; P67PHOX.
DR PRINTS; PR00452; SH3DOMAIN.
DR ProDom; PD000066; SH3; 5.
DR SMART; SM00054; Eph; 2.
DR SMART; SM00027; EH; 2.

```

```

DR SMART; SM00326; SH3; 5.
DR PROSITE; PS00018; EF_HAND; 2.
DR PROSITE; PS50031; EH; 2.
DR PROSITE; PS50002; SH3; 5.
DR PROSITE; PS50002; SH3; 5.
DR Endocytosis; SH3 domain; Repeat; Coiled coil; Calcium-binding.
FT DOMAIN 21 109 EH 1.
FT CA BIND 66 77 EF-HAND 1 (POTENTIAL).
FT DOMAIN 220 309 EH 2.
FT CA BIND 266 277 EF-HAND 2 (POTENTIAL).
FT DOMAIN 325 697 Lys/Leu/Glu/Arg/Gln-Rich (KLERO).
FT DOMAIN 349 691 COILED COIL (POTENTIAL).
FT DOMAIN 732 793 SH3 1.
FT DOMAIN 897 955 SH3 2.
FT DOMAIN 986 1044 SH3 3.
FT DOMAIN 1058 1122 SH3 4.
FT DOMAIN 1139 1198 SH3 5.
SQ SEQUENCE 1270 AA; 143670 MW; EA940C1F6B6A6858 CRC64;

Query Match 80.8%; Score 5063.5; DB 1; Length 1270;
Best Local Similarity 80.6%; Pred. No. 2.5e-166;
Matches 980; Conservative 98; Mismatches 123; Indels 15; Gaps 8;

Qy 1 MAQPTPFGSLDVWVAITVEERAKHQDFLSLKEPIAGFITGDQARNFPFQSGLPQVLAQ 60
Db 1 MAQGTPTFGSLDIWVAITVEERAKHQDFGLKFTAGYITGDQARNFPLQSGLPQVLAQ 60
Qy 61 IVALADNNGRMNOVEFTAMKLIKLGQCYQSPSTLPVPMKQOAPAISA--PARGIG 118
Db 61 IVALADNNDGRMDQLEFSLAMKLIKLGQYPLPSLPSNMLQVAPMAVAAGFQMS 120
Qy 119 GIAMSPPLTAVAPVPMGSIPIVPGMSPPVSVPPAAVPPPLANGAPVYIQLPAPAHPAAT 178
Db 121 GIVGIPPLAANAPVPMISIPVPGMSPPVSV--TVPPLSNGAPAYIQSHPAFAH--SAT 177
Qy 179 WPKSSFSRSGPSQNLTKQAQSFVNASAPPAEAVAPVPOSSRLKTRQLFNSHDKTMSG 238
Db 178 LPKSSFSRSGVAGQINTKQAQSFVNPAPPLVAVAVSSSLKTRQLFNSDQKTMG 237
Qy 239 HLTPQARTIMQSLPQAQASLWNLSDIDQDKLTAEEFTLMLHLIDVAMSGQPLPV 298
Db 238 NLTPQARTIMQSLPQSQSLATWNLSDIDQDKLTAEEFTLMLHLIDVAMSGQPLPV 297
Qy 299 LPPEYIPSPFRVNSGSGMSYSSSVYDQRLPEEPPSEDEQPEPKLVTFEDKKRENF 358
Db 298 LPPEYIPSPFRVNSGSGSLTMSVSVYDQRLPEEPPSEDEQPEPKLVTFEDKKRENF 357
Qy 359 RGSVELEKRGQALLQEKQERLAQLBRAQEKERERQEQAKQOLEKQLEKQEREL 418
Db 358 RGNLELEKRGQALLQEKQERLAQLBRAQEKERERQEQAKQOLEKQEREL 417
Qy 419 ERQEEERKEIEERBAKRELEROROLEWRNRROELNORNEQEGTVVLKARKTLE 478
Db 418 ERQEEERKEIEERBAKRELEROROLEWRNRROELNORNEQEGTVVLKARKTLE 477
Qy 479 FELALNDKQGLQGLQDLIRCRILATQROEIESTNKSRELIATITLQDQLOQESQOMLG 538
Db 478 FELALNDKQGLQGLQDLIRCRILATQROEIESTNKSRELIATITLQDQLOQESQOMLG 537
Qy 539 RLPEKQILSDQLKQVOONSILHDSLLTLKRALFAKELIARQQLREBULDEREFSKQOE 598
Db 538 KPIPEKQILSDQLKQVOONSILHDSLLTLKRALFAKELIARQQLREBULDEREFSKQOE 597
Qy 599 IDVFNNOQLKEIREHSQQLQOKQSLAARLKQEKQEKSELEKQEKEDARQVQERPKQ 658
Db 598 IDVFNNOQLKEIRELYNQQLQOKQDFTEKIKQEKSELEKQEKEDARQVQERPKQ 657
Qy 659 WLEHVQOEQPRPKPEHEDRLKREDSYRKKEAEERAKPEQDQKQSLFHFQEPKATLAT 718
Db 658 WQDRVQOE--RYFPQDEKEKEESQKQEVK--KPEIQEKPKKFHPPEPKGLG 713
Qy 719 QAPWSTTEKGLTISAQSVKVVYRALYPFESSHDEITITQPDIVNVDQSGEPGWL 778
Db 714 QIPMNTEKAPLTIN--QGDVAVVYVYRALYPDASHSDEITTEPGDIIVNVDQSGEPGWL 772

```

QY 779 GGEIKGTGMPANVAEKIPENEPVTPAKVPTDUTSAPAPKLAETREPAVL-PTTSSEPS 837
 DB 773 GGEIKGTGMPANVAEKIPENEPVTPAKVPTDUTSAPAPKLAETREPAVL-PTTSSEPS 828
 QY 838 TTPNNMADPSSSTWSSSNEKPEPTDMNDTMAAOPSLTPSPAGLORSAFTPATATGSSPS 897
 DB 829 TSNMNMADPSSSTWPTNTNDKVESDMDTMAAOPSLTPSPAGLORSAFTPATATGSSPS 888
 QY 898 PVLGGEKEKVEGLQAOLLYPMRACKDHLNFKNSDVTTLVEQDDMMWFGVEQGGKMPFYS 957
 DB 889 PVLGGEKEKVEGLQAOLLYPMRACKDHLNFKNSDVTTLVEQDDMMWFGVEQGGKMPFYS 948
 QY 958 YVKLISGPVRSKSTIDYPTGESPSALSKRVASPAKPAIGEETIATYTESSEQGLTLQ 1017
 DB 949 YVKLISGPVRSKSTIDYPTGESPSALSKRVASPAKPAIGEETIATYTESSEQGLTLQ 1008
 QY 1018 QGDVIVITKKGDMGTGTVDKSGVSPSNVRLKDSGSGTAGKTSGLGKKEPIAIVAS 1077
 DB 1009 QGDVIVITKKGDMGTGTVDKSGVSPSNVRLKDSGSGTAGKTSGLGKKEPIAIVAS 1068
 QY 1078 YVATGEQULTAPGQILIRKKNPGGMBEGELQARQKQIGMPANVYKLSPGTSKIT 1137
 DB 1069 YVATGEQULTAPGQILIRKKNPGGMBEGELQARQKQIGMPANVYKLSPGTSKIT 1128
 QY 1138 PTELEPTAVQPAVCOVIGMVDYTAONDELAESKQIIVLWKEPDMWKGESQGVGF 1197
 DB 1129 PTELEPTAVQPAVCOVIGMVDYTAONDELAESKQIIVLWKEPDMWKGESQGVGF 1188
 QY 1198 PSNYVLTMDPSQ 1213
 DB 1189 PSNYVLTMDPSQ 1204
 QY 1198 PSNYVLTMDPSQ 1213
 DB 1189 PSNYVLTMDPSQ 1204
 RESULT 5
 ID ITN2_HUMAN STANDARD; PRT: 1696 AA.
 AC Q9NZM3; Q95062; Q15812; Q9HAK4; Q9NKE6; Q9NYG0; Q9NZM2; Q9UTL4;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Intersectin 2 (SH3 domain-containing protein 1B) (SH3P18) (SH3P18-like
 WASP associated protein).
 GN ITSN2 OR SH3D1B OR SWAP OR KIAA1256.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP TISSUE=Brain; (ISOFORMS 1 AND 3).
 RC TISSUE=Brain; (ISOFORMS 1 AND 3).
 RX MEDLINE=20382722; PubMed=10922467;
 RA Pucharcos C., Estivill X., de la Luna S.;
 RA "Intersectin 2, a new multimodular protein involved in clathrin-
 mediated endocytosis";
 RT FEBS Lett. 478:43-51(2000).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RC Seifert M., Engel M., Welter C.;
 RA "Intersectin 2 (SH3D1B), human homolog of mouse Eez protein";
 RT Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RC TISSUE=Brain;
 RX MEDLINE=20039619; PubMed=10574462;
 RA Nagase T., Ishikawa K.-I., Kikuno R., Hirose M., Nomura N.,
 RA Ohara O.;
 RA "Prediction of the coding sequences of unidentified human genes. XV.
 RT The complete sequences of 100 new cDNA clones from brain which code
 RT for large proteins in vitro";
 RL DNA Res. 6:337-345(1999).
 RN [4]
 RP SEQUENCE OF 258-720 FROM N.A. (ISOFORM 2), AND SEQUENCE OF 1261-1696

RP FROM N.A.
 RC TISSUE=Embryo;
 RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
 RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
 RA Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,
 RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,
 RA Makamatsu A., Nakamura Y., Nagahara K., Masuno Y., Oshima A.;
 RT "NEO human cDNA sequencing project";
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE OF 821-1284 FROM N.A. (ISOFORM 1).
 RA Wilson L.A., Fields D., Cruz L., Friesen J., Simionovitch K.A.;
 RT "SH3P18-like wasp associated protein (SWAP): A multiple SH3 domain
 RT containing protein that interacts with WASP";
 RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RP SEQUENCE OF 944-1191 FROM N.A. (ISOFORM 4).
 RC TISSUE=Bone marrow;
 RX MEDLINE=98294438; PubMed=9630982;
 RA Sparks A.B., Hoffman N.G., McConnell S.J., Fowles D.M., Kay B.K.;
 RT "Cloning of ligand targets: systematic isolation of SH3 domain-
 RT containing proteins";
 RL Nat. Biotechnol. 14:741-744(1996).
 RN [7]
 RP ALTERNATIVE SPLICING.
 RC TISSUE=Brain, and Fetal liver;
 RX MEDLINE=21548828; PubMed=11690630;
 RA Pucharcos C., Casas C., Nadal M., Estivill X., de la Luna S.;
 RT "The human intersectin genes and their spliced variants are
 RT differentially expressed";
 RL Biochim. Biophys. Acta 1521:1-11(2001).
 CC -1- FUNCTION: Adapter protein that may provide indirect link between
 CC the endocytic membrane traffic and the actin assembly machinery.
 CC May regulate the formation of clathrin-coated vesicles.
 CC -1- SUBUNIT: Belongs to a complex that may contain multimers of ITSN1,
 CC ITSN2 and Eps15, and different partners according to the step in
 CC the endocytic process.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=4;
 CC Comment=Additional isoforms seem to exist;
 CC Name=1; Synonyms=ITSN2-L;
 CC IsoId=Q9NZM3-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q9NZM3-2; Sequence=VSP_003892;
 CC Name=3; Synonyms=ITSN2-S1;
 CC IsoId=Q9NZM3-3; Sequence=VSP_003893, VSP_003894;
 CC Name=4; Synonyms=ITSN2-S2, SH3P18;
 CC IsoId=Q9NZM3-4; Sequence=VSP_003895;
 CC -1- TISSUE SPECIFICITY: Ubiquitous. Isoform 1 is primarily expressed
 CC in adult heart and liver.
 CC -1- MISCELLANEOUS: Overexpression results in the inhibition of the
 CC transferrin uptake and the blockage of the clathrin-mediated
 CC endocytosis.
 CC -1- SIMILARITY: Contains 1 C2 domain.
 CC -1- SIMILARITY: Contains 1 DBL-homology (DH) domain.
 CC -1- SIMILARITY: Contains 1 EF-hand calcium-binding domain.
 CC -1- SIMILARITY: Contains 1 PH domain.
 CC -1- SIMILARITY: Contains 5 SH3 domains.
 CC -1- SIMILARITY: Contains 2 EH domains.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC EMBL; AF182198; AAF59903.1; -;
 CC EMBL; AF182199; AAF59904.1; -;
 CC EMBL; AF248540; AAF63600.1; ALT INIT.
 CC EMBL; AB033082; BAA86570.1; ALT INIT.

DR EMBL; AK021545; BAB13841.1; -
 DR EMBL; AK000302; BA91068.1; -
 DR EMBL; AF001630; AAD00899.1; -
 DR EMBL; U61167; AAC50593.1; -
 DR HSSP; P29355; 1SEM.
 DR Genew; HGNC:6184; ITS2.
 DR MIM; 604464; -
 DR GO; GO:0005070; F:SH3/SH2 adaptor protein activity; TAS.
 DR InterPro; IPR000008; C2.
 DR InterPro; IPR002048; EF-hand.
 DR InterPro; IPR000261; EPS15_homology.
 DR InterPro; IPR001849; PH.
 DR InterPro; IPR001452; SH3.
 DR Pfam; PF00168; C2; 1.
 DR Pfam; PF00036; ehand; 2.
 DR Pfam; PF00169; PH; 1.
 DR Pfam; PF00621; RhogEF; 1.
 DR Pfam; PF00018; SH3; 5.
 DR PRINTS; PR00452; SH3DOMAIN.
 DR ProDom; PD000066; SH3; 5.
 DR SMART; SM00239; C2; 1.
 DR SMART; SM0054; EFh; 2.
 DR SMART; SM0027; EH; 2.
 DR SMART; SM00233; PH; 1.
 DR SMART; SM00325; RhogEF; 1.
 DR SMART; SM00326; SH3; 5.
 DR PROSITE; PS00004; C2_DOMAIN_2; 1.
 DR PROSITE; PS00010; EH_2; 1.
 DR PROSITE; PS00018; EF-hand; 1.
 DR PROSITE; PS00031; EH; 2.
 DR PROSITE; PS00003; PH_DOMAIN; 1.
 DR PROSITE; PS00002; SH3; 5.
 DR EndoCytoS; SH3 domain; Repeat; Coiled coil; Calcium-binding;
 KW Alternative splicing.
 FT DOMAIN 21 109 EH 1.
 FT CA BIND 66 78 EF-HAND (POTENTIAL).
 FT DOMAIN 243 332 EH 2.
 FT DOMAIN 360 755 COILED COIL (POTENTIAL).
 FT DOMAIN 756 817 SH3 1.
 FT DOMAIN 897 955 SH3 2.
 FT DOMAIN 980 1038 SH3 3.
 FT DOMAIN 1052 1116 SH3 4.
 FT DOMAIN 1126 1185 SH3 5.
 FT DOMAIN 1208 1394 DH.
 FT DOMAIN 1433 1543 PH.
 FT DOMAIN 1555 1651 C2 DOMAIN.
 FT VARSPLIC 621 647 Missing (in isoform 2).
 FT VARSPLIC 1234 1248 Missing (in isoform 3).
 FT VARSPLIC 1249 1696 Missing (in isoform 3).
 FT VARSPLIC 1192 1696 Missing (in isoform 4).
 FT CONFLICT 290 290 V->I (IN REF. 4).
 FT CONFLICT 678 678 R->G (IN REF. 4).
 FT CONFLICT 827 827 KAVSPK->FAAAT (IN REF. 5).
 FT CONFLICT 944 950 WEPKSYV->EFAAAT (IN REF. 6).
 FT CONFLICT 1278 1284 GEKMPVO->VDAANS (IN REF. 5).
 FT CONFLICT 1552 1552 K->Q (IN REF. 2 AND 3).
 SQ SEQUENCE 1696 AA; 193329 MW; 17B5C8629BECF9B CRC64;

Query Match 48.9%; Score 3068; DB 1; Length 1696;
 Best Local Similarity 51.0%; Pred. No. 3.8e-110;
 Matches 656; Conservative 187; Mismatches 274; Indels 170; Gaps 30;

Qy 1 MAQPTTPGSGLDVWATVEBAKDDQFLSKPLAGITGDQARNFFGSLPQVLAQ 60
 Db 1 MAQPTTPANGGPNNWATITSEERTKDRQFDNLKPSGYITGDQARNFPLQSGLPAPVLA 60
 Qy 61 IWLADNMNDGRMQRVERSIAMKLIKLTQGYQLPSTLPVWKQGPV--ATSSAPAFGIG 118

Db 61 IWLADNMNDGRMQRVERSIAMKLIKLTQGYQLPVLPTIMKQPFMFSLISA-RGOMG 119
 Qy 119 GIASMPPLTAVAPVPMG-----SIPVNGSPPLVSVPPAAVPLANGAPV 165
 Db 120 ---SMPNISIPQPLPAPAPITSLSSATSGTNLPPLMPTLPVSVSTSLP---NGTASL 173
 Qy 166 IQPLPAHAHAATMPKSSSFS--RSGGSGQNTLTKQAQS-FDVASAP----- 211
 Db 174 IQPLP-IPYSSSTIPHSSSYSLMMGFG---GASIQAAQSLIDGSSSSSTSLSGNS 229
 Qy 212 ----AAEMAVPQSSRLKYROLFNSHDXMTSGHLTGPAARTILMQSSLPOAQLASINLSD 267
 Db 230 PKGTSEMAVQPFRLTKYRQKFNLTDXSMGYSLSGFARNALLQNSLSQTLATITLAD 289
 Qy 268 IDQGKLTAEFFILAMHLDVANSGLPPLVPEYIPPSFRARRSSGSGSVSSSDVQ 327
 Db 290 VDGSGQKABEFFILAMHLDMAKAGPLPLTLPPELVPSPFR---GKQI---DSING 341
 Qy 328 RLPEPSEDEQPEKPLPTFEDKKRNFERSGVELEKRRQALLJEOQRKEERTLAQLER 387
 Db 342 TLPSYQMG--EEFQKLPPTFEDKKRANYERGMLEKRRQALMEQQQRAEKAQKEK 400
 Qy 388 AEQKERERQEQAKRQLELEKQLEKQRELEQREERKEIRREBAKRELEQROLE 447
 Db 401 EEMERKQRELEQEWKQLELEKREKQRELEQREERERDRIRREBAKRELEQROLE 460
 Qy 448 WERNRQELLNQRKKEQGYVYLKARKTLEFELEALNDKKHQLQEGLODIRCLATQRO 507
 Db 461 WERIRQELLNQRKKEQGYVYLRNLSKKKNLHLEALNGHQQISGLQVRLKKQYQKT 520
 Qy 508 EIESTNKSRELRALAEITHLQOQLOESQOMGLRLEPEKQILSDLOKQVQ-ONSLEHDSILT 566
 Db 521 ELEVLQOCDEIWEIQOQLOELEYONKLIYVPEKQLANERIKNNQFNTSPSGVSL 580
 Qy 567 LKRALKEKELARQOLREQLDEVERETRSKLOEIDVFNNQ----- 605
 Db 581 HKSLKEKEELC-QRLKQDLALBKETASKLSMDSFNNQKCGMDDSVLQCLLSLSC 639
 Qy 606 ----LKELEHNSKQOLQKNSLEAARLKQEQEKSLLEKQK---EDAQRQER 655
 Db 640 NNLFLLKEKRETNTOQLALEQLYKIKRDKLEIKERKLELMQKLEDEAARAKAQGK 699
 Qy 656 DKQWLEHVQEQEQRPRKPHBEDLTKREDSVRKKEABERAKPEMQDQKSLRFFHQBP 715
 Db 700 ENLWKENLKRKEE-----EKQKLEQEKQEKQKEERAKQEKQKDXD----- 743
 Qy 716 LATQAPWSTTEKGLTISAQESVK-----VYYRALYPFESRSHDEITIQGDIWVDES 770
 Db 744 -----TLKAEKKRETSVLYNTRALVPEARNDHDSFNSGDI IQDEK 788
 Qy 771 QTGEPGULGELKGTGMPANTAEKLP--ENEVPTAKPVTDLTSPAPKLAIRETPAP 828
 Db 789 TVGEPGLYSGFQGNFPMPCNVYKMPSENE-----KAVSPKAL----- 830
 Qy 829 LPTVSSPSTTPNNWADPSSSWPSSNKEPDTOWDWTAAQPSLTVSAGOLRORSAFTP 888
 Db 831 LPTVSLASATG-----TSSEPLSSNQPASVTDYQ--VSPSNLTWTSMQ--KKSAPTR 880
 Qy 889 ATATGSPSPVLYGGEKVEYEQALQALYPMRAKDNHNLFNKSDVITYLEQDMMWFEVQ 948
 Db 881 TVSPG-SVSPHGGQVVENLKQALCSMTAKDNHNLNFSGHDITYLEQDMMWFEVH 939
 Qy 949 GQKGMFPKSYVKLISGVRKSTSIDTPTSPSLKVAS--PAKALPGEETIAMTYT 1006
 Db 940 GGRGMFPKSYVKLIPGSEVK-----REEPALVAANVKKPSTSAYSVGEETIAYLPY 991
 Qy 1007 ESSQEGDLTQGGQVIVYTKDGDWMTGTGDSGVSPSNVYRKDSGSGTAKTSGISG 1066
 Db 992 SSVEPGDLTTEGSEILVTQDGEWMTGSGIDRSGTIPSNVYRKDDESFSASKASGASN 1051
 Qy 1067 KKPEIAQVIAVYATGEGQTLTLAGQILIRKKNPGCMWEGELQARKKQIQIGFPANVY 1126

Db 1052 KPEIAQVTSAYVAGSEQLSLAPQGLILKNTSGMGELOARCKRKQKWPESHV 1111
 QY 1127 KLLSPGTSKITPTELPEKTAQVAVGVGMVDTAQNDELAFSGQIINLVNKEPDMW 1186
 Db 1112 KLLGSSSEATATAFH-----VCCVIMYDPAANNEDELSFGQLINVKDDPDW 1164
 QY 1187 KGEVSGQVGLFPSNVYKLTITMDPSQ 1213
 Db 1165 QGEINGVTGLFPSNVYKMTTSDPSQ 1191

RESULT 6
 ITN2 MOUSE
 ID ITN2 MOUSE STANDARD; PRT; 1658 AA.
 AC Q9ZOR6; Q9ZOR5;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Intersectin 2 (SH3 domain-containing protein 1B) (EH and SH3 domains
 protein 2) (EH domain and SH3 domain regulator of endocytosis 2).
 GN ITSN2 OR SH3D1B OR ESE2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
 RA MEDLINE=99164083; PubMed=10064583;
 RA Sengar A.S., Wang W., Bishay J., Cohen S., Egan S.B.;
 RT "The EH and SH3 domain Ese proteins regulate endocytosis by linking to
 RT dynamin and Eps15.";
 RL EMBO J. 18:1159-1171(1999).
 CC -1- FUNCTION: Adapter protein that may provide indirect link between
 CC the endocytic membrane traffic and the actin assembly machinery.
 CC May regulate the formation of clathrin-coated vesicles.
 CC -1- SUBUNIT: Belongs to a complex that may contain multimers of ITSN1,
 CC ITSN2 and Eps15, and different partners according to the step in
 CC the endocytic process.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=1; Synonyms=Ese2L, Long;
 CC IsoId=Q9ZOR6-1; Sequence=Displayed;
 CC Name=2; Synonyms=Ese2, Short;
 CC IsoId=Q9ZOR6-2; Sequence=VSP_003896, VSP_003897;
 CC -1- TISSUE SPECIFICITY: Widely expressed in adult tissues.
 CC -1- DEVELOPMENTAL STAGE: Widely distributed throughout the adult
 CC forebrain. Prominent expression was observed in the neocortex, the
 CC piriform cortex, the pyramidal cell layers of hippocampus, the
 CC dentate gyrus, in several nuclei of the thalamus and hypothalamus
 CC and in the amygdala.
 CC -1- SIMILARITY: Contains 2 C2 domains.
 CC -1- SIMILARITY: Contains 1 DBL-homology (DH) domain.
 CC -1- SIMILARITY: Contains 1 EF-hand calcium-binding domain.
 CC -1- SIMILARITY: Contains 1 PH domain.
 CC -1- SIMILARITY: Contains 5 SH3 domains.
 CC -1- SIMILARITY: Contains 2 EH domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL; AF132479; AAD19747.1; -;
 CC EMBL; AF132480; AAD19748.1; -;
 CC HSSP; P29355; 1SEM.
 CC MGI; MGI:1358049; SH3D1B.
 CC InterPro; IPR000008; C2.
 CC InterPro; IPR002048; EF-hand.
 CC InterPro; IPR000261; Eps15_homology.

DR InterPro; IPR000108; Neu_cyt_fact_2.
 DR InterPro; IPR001849; PH.
 DR InterPro; IPR000219; RhogEF.
 DR InterPro; IPR001452; SH3.
 DR Pfam; PF00168; C2; 1.
 DR Pfam; PF00036; ehfand; 2.
 DR Pfam; PF00169; PH; 1.
 DR Pfam; PF00621; RhogEF; 1.
 DR Pfam; PF00018; SH3; 5.
 DR PRINTS; PR00459; P67PHOX.
 DR ProDom; PD000066; SH3; 5.
 DR SMART; SM00239; C2; 1.
 DR SMART; SM00054; EFh; 2.
 DR SMART; SM00027; EH; 2.
 DR SMART; SM00233; PH; 1.
 DR SMART; SM00325; RhogEF; 1.
 DR SMART; SM00326; SH3; 5.
 DR PROSITE; PS50004; C2 DOMAIN_2; 1.
 DR PROSITE; PS50010; DH_2; 1.
 DR PROSITE; PS00018; EF_HAND; 1.
 DR PROSITE; PS50031; EH_2.
 DR PROSITE; PS50003; PH_DOMAIN; 1.
 DR PROSITE; PS50002; SH3; 5.
 KW Endocytosis; SH3 domain; Repeat; Coiled coil; Calcium-binding;
 KW Alternative splicing.
 FT DOMAIN 21 109
 FT CA_BIND 66 78
 FT DOMAIN 244 333
 FT DOMAIN 364 716
 FT DOMAIN 717 778
 FT DOMAIN 851 909
 FT DOMAIN 941 999
 FT DOMAIN 1013 1077
 FT DOMAIN 1087 1146
 FT DOMAIN 1169 1356
 FT DOMAIN 1395 1505
 FT DOMAIN 1517 1613
 FT VARSPLIC 1187 1197
 FT C2 DOMAIN.
 FT DDLQVIEVFQ -> GLQLFEQKTL (in isoform
 2).
 FT VARSPLIC 1198 1658
 FT VARSPLIC 1198 1658
 FT CONFLICT 451 452
 FT CONFLICT 1161 1161
 FT SEQUENCE 1658 AA; 188776 MW; 7050EFC5F7983A5 CRC64;
 SQ
 Query Match 48.6%; Score 3049.5; DB 1; Length 1658;
 Best Local Similarity 52.4%; Pred. No. 1.9e-109;
 Matches 658; Conservative 177; Mismatches 275; Indels 145; Gaps 32;
 QY 1 MAQFTPPGSLDVAITVEERAKHQQLSLKPIAGFITGDQANFFPQGLPQVLAQ 60
 Db 1 MAQFTPANQGPNNWAIETSEERTGDKQFDMLKESGQYITGDQARTFLQGLPAPVLA 60
 QY 61 IVALADNNDRMOVERSIMKLIKTLQGYVLPSTLPYPMKQPV--AASSAPAFGIG 118
 Db 61 IVALSLDKDKMDQEFISIMKLIKTLQGYVLPVLPYPMKQPVMPSPISIA-RFGWG 119
 QY 119 GIASNP-----PLTAVPV--PMG-----SIPVGSPLVSSVPPAAVPLANGAP 164
 Db 120 ---SNPNSTHQPLPVPAIPATPLSSATSGSIPPLMAPLVBVSSTSLP---NGTAS 173
 QY 165 VIQPLPAFPAHPATWPKSSFS--RSGFQSLNTKLQKAGS-FDVASAP----- 211
 Db 174 LIQPL-SIPYSSSTLPHASSYSILMMGGRG---GASIQAGSLIDGSSSSSTSTASLGN 229
 QY 212 -----AARNAVPOSSRLKYROLFNSHDKTNGHLTGPARTIMOSSLPQAGLASIMWLS 266
 Db 230 SPKTGTSMAVPOBPRLKRYRQFNSLDKMGSGYLSGFAVALQSNLSQTQLATITWLA 289
 QY 267 DIDDGKLTAAEFILAMLIDVAMSGQLPVLPPEYVPPSFRRRSGSGMSVSSSSVD 326
 Db 290 DIDDGKLTAAEFILAMLIDVAMSGQLPVLPPEYVPPSFR-----GKQV---DSVN 341

```

327 QRLPEBSSBDEQPEKLPYTFEDKKRNEPERSVLEKKRQALLQEQKREQLQLE 386
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
342 GTLPSYQKQ-EEBQKQLPYTFEDKKRNEPERSVLEKKRQALLQEQKREQLQLE 400
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
387 RAEQERKERQEQEAKQLLEKQLEKQLEKQLEKQLEKQLEKQLEKQLEKQLEKQLE 446
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
401 KEERERKQREQEQEAKQLLEKQLEKQLEKQLEKQLEKQLEKQLEKQLEKQLEKQLE 460
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
447 EWEENRQELINQNKQEGTVLKKRKTLEFEELEALNDKKHLEGLQDICKRLATQR 506
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
461 EWERLRQELINQNKQEGTVLKKRKTLEFEELEALNDKKHLEGLQDICKRLATQR 520
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
507 QEISTNSRRLRLAETHLQOQLOESQOQMLGRLLPEKQILLSDQKQOQNSLHRDLSLT 566
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
521 TELEVLQDQOQLETHLQOQLOESQOQMLGRLLPEKQILLSDQKQOQNSLHRDLSLT 580
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
567 LKRALLEAKELARQOQLEQDEVERETRSKLOEIDVFNQQLKELEHSHKQOQLOKQSLIA 626
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
581 LHKSSSEKELQOQLEKQLEKQLEKQLEKQLEKQLEKQLEKQLEKQLEKQLEKQLE 640
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
627 ARLKQEQERKSLF-LEKQK--EDAQRRVQERDKQMLEHVQOEQOPRRKPHEDRLKR 682
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
641 KRDKLKEIERKRLQEQLEKQLEKQLEKQLEKQLEKQLEKQLEKQLEKQLEKQLEKQLE 694
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
663 EDSVRKKEAERAKEMQDKQSLFPHHQEPAKLATQAPMSTETEGPLTISAQESVKYVY 742
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
695 EKSQDKTEER--KAEAKQSE-----TAAH-----LVN 721
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
743 YRALYFPESRSHDEITIQGDIVWVDESQGEPCWMLGELKQGTMPFANYAEKIPENEV 802
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
722 YRALYFPFARHNDHESFSSGDIQVDEKTEVEPGWLYSFGQKPFMFCVNYEKLSE- 780
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
803 PTPAKPVTDLTGAPAKLALRETPALPVTSESEPTTNMADFESTPSSNEKPEETDN 862
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
781 -----KALSPKXALLPPTVSLSATSTS--SQPASVTDYHNV--SFSNLTVNT-- 824
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
863 WDTMAAQSLTVPSAGQURQSPAFATATSSPSPVGGQGVKGLQAOALPYPARAKD 922
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
825 --TW-----QKSAFRTVSPG-SVSPHGGQGVNLAQALCSWTAKKE 867
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
923 NNLNFKSDVITVLEEQDMMWFGVQKGMFPKPSYVKLISG-PYRKS--TSIDTPTES 979
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
868 NNLNFKSDVITVLEEQDMMWFGVQKGMFPKPSYVKLIPENEQREPEALYAAVYKK 927
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
960 PASLRVASPAKPAIP-GEFTIANYTESSEQGLTQOQDVIVYTKKQDMMWGTGVD 1038
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
928 PTS--TAYPTSTAYPVGEDYIALYSYSVPEPGLTFEGEELIVTQKQDGMWMTGSIGE 984
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1039 KSGVPSNYYVRLKDESGSTAGKTGSLGKPEIAQVIAVSATGEOLTLIAPGQILIRK 1098
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
965 RTGIPSPNYYVRKDDENFGNASKSASNNKPEIAQVIAVSATGEOLTLIAPGQILIRK 1044
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1099 KNPQGMWGELOARAKKEQIQGMFPANYVLLSPGTSKITPTTELPTAVQAPVQVIGMYD 1158
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1045 KNTSGMWGELOARAKKEQIQGMFPANYVLLSPGTSKITPTTELPTAVQAPVQVIGMYD 1097
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1159 YTAQNDDELAFSKGQIINVLANKEDDPMWKGVSQGVGLFPSSNYVKLTDMDSQO 1213
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1098 YWANNEDDELAFSKGQIINVLANKEDDPMWKGVSQGVGLFPSSNYVKLTDMDSQO 1152
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
CX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Melanoma;
RX MEDLINE=94239734; PubMed=8183552;
RA Wong M.T., Kraus M.H., Carlomagno F., Zelano A., Druck T.,
RA Croce C.M., Huebner K., di Fiore P.P.,
RT "The human eps15 gene, encoding a tyrosine kinase substrate, is
RT conserved in evolution and maps to 1p31-p32."
RL Oncogene 9:1591-1597(1994).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94181254; PubMed=8134107;
RA Bernard O.A., Mauchault M., Mecucci C., van den Berghe H.,
RA Berger R.;
RT "A novel gene, AF-1p, fused to HRX in t(1;11)(p32;q23), is not
RT related to AF-4, AF-9 nor ENL."
RL Oncogene 9:1039-1045(1994).
RN [3]
RP INTERACTION WITH CRK.
RX PubMed=8662907;
RA Matsuda M., Ota S., Tanimura R., Nakamura H., Matuoka K., Takenawa T.,
RA Nagashima K., Kurata T.;
RT "Interaction between the amino-terminal SH3 domain of CRK and its
RT natural target proteins."
RL J. Biol. Chem. 271:14468-14472(1996).
RN [4]
RP STRUCTURE BY NMR OF 217-311.
RX MEDLINE=20222955; PubMed=10757979;
RA Emmon J.L., de Beer T., Overduin M.;
RT "Solution structure of Eps15's third EH domain reveals coincident
RT Phe-Tyr and Asn-Pro-Phe binding sites."
RL Biochemistry 39:4309-4319(2000).
RN [5]
RP STRUCTURE BY NMR OF 121-215.
RX MEDLINE=98387926; PubMed=9721102;
RA de Beer T., Carter R.E., Lobel-Rice K.E., Sorokin A., Overduin M.;
RT "Structure and Asn-Pro-Phe binding pocket of the Eps15 homology
RT domain."
RL Science 281:1357-1360(1998).
CC -1- FUNCTION: Involved in cell growth regulation. May be involved in
CC the regulation of mitogenic signals and control of cell
CC proliferation. Involved in the internalization of ligand-inducible
CC receptors of the receptor tyrosine kinase (RTK) type. In
CC particular EGFR (By similarity).
CC -1- SUBUNIT: Interacts with Stomatin 2. Interacts with CRK via its SH3-
CC binding sites.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic. Recruited to the plasma
CC membrane upon EGFR activation and localizes to coated pits (By
CC similarity).
CC -1- TISSUE SPECIFICITY: Ubiquitously expressed.
CC -1- PTM: Phosphorylation on Tyr-849 is involved in the internalization
CC of EGFR. Not required for membrane translocation after EGF
CC treatment or for targeting to coated pits, but essential for a
CC subsequent step in EGFR endocytosis (By similarity).
CC -1- DISEASE: Involved in a t(1;11)(p32;q23) chromosomal translocation
CC in acute leukemias causing fusion to the trithorax (MLL or HRX)
CC gene product which contains DNA-binding motifs resulting in a
CC rogue activator protein.
CC -1- SIMILARITY: Contains 2 EF-hand calcium-binding domains.
CC -1- SIMILARITY: Contains 2 ubiquitin-interacting motif (UIM) repeat.
CC -1- DATABASE: NAME=Atlas Gene. Cytogenet. Oncol. Haematol.;
CC WWW="http://www.intoblogen.fr/services/chromosome/Genes/AF1pD11.html".
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/

```

```

CC or send an email to license@slb.ch).
CC EMBL; U0707; AAA52101.1; -.
DR EMBL; Z29064; CAA82305.1; -.
DR PIR; S43074; S43074.
DR PDB; 1C07; 19-JUL-00.
DR PDB; 1EH2; 22-JUL-99.
DR PDB; 1F8H; 01-NOV-00.
DR PDB; 1FEL; 01-NOV-00.
DR Geneb; HGNC:3419; EP815.
DR MIM; 600051; -.
DR GO; GO:0008283; P:cell proliferation; TAS.
DR GO; GO:0007173; P:EGF receptor signaling pathway; TAS.
DR GO; GO:0007048; P:oncogenesis; TAS.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR00261; EPS15_homology.
DR InterPro; IPR003903; UIM.
DR Pfam; PF00036; ehfand; 4.
DR SMART; SM00054; Eph; 3.
DR SMART; SM00027; EH; 3.
DR SMART; SM00726; UIM; 3.
DR PROSITE; PS00018; EF_HAND; 2.
DR PROSITE; PS00031; EH; 3.
DR PROSITE; PS03030; UIM; 2.
DR SH3-binding; Phosphorylation; Calcium-binding; Repeat;
KW Chromosomal translocation; Proto-oncogene; 3d-structure.
FT DOMAIN 15 104 EH 1.
FT CA_BIND 128 216 EH 2.
FT CA_BIND 173 184 EH 3.
FT CA_BIND 224 314 EF_HAND 1 (POTENTIAL).
FT CA_BIND 236 247 EF_HAND 2 (POTENTIAL).
FT DOMAIN 599 827 EF_HAND 2 (POTENTIAL).
FT REPEAT 599 601 15 X 3 AA REPEATS OF D-P-F.
FT REPEAT 623 625 1.
FT REPEAT 629 631 2.
FT REPEAT 631 631 3.
FT REPEAT 634 636 4.
FT REPEAT 640 642 5.
FT REPEAT 645 647 6.
FT REPEAT 651 653 7.
FT REPEAT 664 666 8.
FT REPEAT 672 674 9.
FT REPEAT 692 694 10.
FT REPEAT 709 711 11.
FT REPEAT 737 739 12.
FT REPEAT 798 800 13.
FT REPEAT 804 806 14.
FT REPEAT 825 827 15.
FT DOMAIN 851 870 UIM 1.
FT DOMAIN 877 896 UIM 2.
FT DOMAIN 768 850 PRO-RICH.
FT SITE 768 774 SH3-BINDING.
FT MOD_RES 849 849 PHOSPHORYLATION (BY EGFR) (BY
FT CONFLICT 822 822 SIMILARITY).
FT SEQUENCE 896 AA; 38673 MW; A1B9FB04A07FABEB CRC64;

```

```

Db 216 KT-----WVSPAKAKYDIEIFLKTDXMDMGFSGEVEIREIFLKTGL 257
Qy 255 POAQLASINWLSIDIDODGKLTAEFLIAMLHLDVAM-SQGPUPVUPPEYIPPSFRVMS 313
Db 258 PSTLAAHWISLCTDXTDCKSLSDQFALAFHLISQXIKIGIDPPHYLTPEMIPPS---DFA 314
Qy 314 GSGMSVYSSSVYDQRLPEEPPSSDEQPEKKLPVPEDEKKRNFERSGVLEKRRQALTE 373
Db 315 SLQKNTIGSSPV-----ADFSALK-----ELDTLNNETVD 344
Qy 374 QQRKEQERLAEERAEQERKERERQEAQKOLEKOLEKORERLORERERKEIERR 433
Db 345 LQREK-----NNVQDLKEK-----DTIKQRTSVQ---DLQDVEQHE 380
Qy 434 EAAKRELERQRLERERNRQELINDRNKEQEGTVLKKARKRTLEBELANDKXQQLG 493
Db 381 NTNLOKLOAKOQ-----QVEELIDE-----LDEQKQLEB 410
Qy 494 KLQDRCRATQROEIESTNKSRELRIAEITLLOOQLQESQOMGLRLEKQILSDQLK 553
Db 411 QLEKVAKKCAEEAQLISSLK-----AELTQESQISTYEELAK-----ARELSR 456
Qy 554 VOQNSLHRDLSLTLLKRALEAKELARQOLREQLDEVERETRS-----KLQEIDVFNNOJK 607
Db 457 LQQETAE-----LEESVSGKAQLEPLQOHQDSQQLISSMQMKLEMKDLENHNSQLN 510
Qy 608 ELREIHS-----KQOLQKORSLEAPLAKQEKQERKSLEI--- 641
Db 511 WCSSPHSILVNGATDYCSLSTSSSEPNANLNEHVEGCSNLESEPIHOESPARSGPELLPSG 570
Qy 642 EKQKEDAQRRVQERDQWLEHVQOEQRPPRPKPHEDRLKREDSVAKKAEE- 644
Db 571 VTDENEVTTAVTEKQSELDN-----NRSKEDPPNNVSSSLTGPVADNLDLFPQS 622
Qy 695 -----AKPEMOKOSRLFFHPOEPAK-----LATQAPWSTTEKGLTISAQESV 738
Db 623 DPFVSGDPFKDPPFGKIDPFGDPPFGSDPPFASDCFRQSTDPFASSIDPPFAANNSI 682
Qy 739 KVVYRYRALYPPESRSHDEITTPGDIWMVDESQT-----GEPGMLG- 780
Db 683 TSV-----ETLKINDPFAFGVVAASDSDATPFAVFNESGCGFADFSTLSKYN 734
Qy 781 -----ELKGTGWFPAN-----YAEKIPENEVTPAKPVTDLTSAPAKLARETPAPLP 830
Db 735 NEDPFESATSSSVSNVITKNVFETSVKSEDEPPALP-----PRIGTPTRCPPLP 785
Qy 831 -----VTSSEP-----STTPNNMADFSTWPSNENKPEETDNWPT 865
Db 786 PKRSTINKLDSPPFDLNDPPQFPFONDSPKREKDPFCDPFTSATTTTNKADPSPNPN 845
Qy 866 WAAQPS 871
Db 846 PSAYPS 851

```

```

Query Match 7.8%; Score 486.5; DB 1; Length 896;
Best Local Similarity 20.5%; Pred. No. 4,7e-12;
Matches 198; Conservative 150; Mismatches 273; Indels 345; Gaps 29;

```

```

Qy 15 MAITVERAKHQDQFLSLKPIAGFITGDQARNFPQSGLPQVLAQIMALADWNNNGRMD 74
Db 122 MAVKPEPDKAKYDAIFDLSFVNGFSLGDKYKPYLANSKLPVLDLGRVWELSDIDHGMID 181
Qy 75 QVEFSIAMKLIKLKIGYOLPSTLPPVMQGPVAISSAPAFGIGIASMPPLTAVAPVP 134
Db 182 RDEFAVAMFLVYCALB-----KEPVPM 203
Qy 135 GASIPVVGMSPLVSSVPPAAPPPLANGAPPVIOPLPAFAHPATWPKSSFSKSGGSQL 194
Db 204 -----SLPPALVPP-----SKR 215
Qy 195 NTQLQKAGSFDVASADPPAAEMAVPQSSRLKYROLFNSHDKTMSGHLTGPOARTILMOSSI 254

```

```

RESULT 8
ID EP15_MOUSE STANDARD; PRT; 897 AA.
AC P42567;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Epidermal growth factor receptor substrate 15 (protein Epals) (AF-1p
DE protein).
GN EPS15.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=fibroblast;

```

RX MEDLINE=93361014; PubMed=7689153;
 RA Paoletti F., Minichiello L., Matochkova B., Wong W.T., di Fiore P.P.;
 RT "Eps15, a novel tyrosine kinase substrate, exhibits transforming
 RT activity.";
 RL Mol. Cell. Biol. 13:5814-5828(1993).
 RN (2)
 RP PHOSPHORYLATION OF TYR-850, AND MUTAGENESIS OF TYR-850.
 RX MEDLINE=20411232; PubMed=10953014;
 RA Confalonieri S., Salcini A.E., Puri C., Tacchetti C., Di Fiore P.P.;
 RT "Tyrosine phosphorylation of Eps15 is required for ligand-regulated,
 RT but not constitutive, endocytosis.";
 RL J. Cell Biol. 150:905-912(2000).
 RN [3]
 RP STRUCTURE BY NMR OF 7-105.
 RX MEDLINE=99400466; PubMed=10471276;
 RA Whitehead B., Tessari M., Carotenuto A.,
 RA van Bergen en Henegouwen P.M., Vulliamis G.W.;
 RT "The Eps1 domain of Eps15 is structurally classified as a member of
 RT the SH3 domain class of EF-hand-containing proteins.";
 RL Biochemistry 38:11271-11277(1999).
 CC -1- FUNCTION: Involved in cell growth regulation. May be involved in
 CC the regulation of mitogenic signals and control of cell
 CC proliferation. Involved in the internalization of ligand-inducible
 CC receptors of the receptor tyrosine kinase (RTK) type, in
 CC particular EGFR.
 CC -1- SUBUNIT: Interacts with CRK via its SH3-binding sites (by
 CC similarity). Interacts with Steen 2.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic. Recruited to the plasma
 CC membrane upon EGFR activation and localized to coated pits.
 CC -1- PTM: Phosphorylation on Tyr-850 is involved in the internalization
 CC of EGFR. Not required for membrane translocation after EGF
 CC treatment or for targeting to coated pits, but essential for a
 CC subsequent step in EGFR endocytosis.
 CC -1- SIMILARITY: Contains 3 EF-hand domains.
 CC -1- SIMILARITY: Contains 2 EF-hand calcium-binding domains.
 CC -1- SIMILARITY: Contains 2 ubiquitin-interacting motif (UIM) repeats.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL; L21768; AAA02912.1; .
 DR PIR; A54696; A54696.
 DR PDB; 1KYF; 12-JUN-02.
 DR PDB; 1KYU; 12-JUN-02.
 DR PDB; 1QUT; 11-SEP-02.
 DR MCD; MGI:104583; Eps15.
 DR InterPro; IPR002048; EF-hand.
 DR InterPro; IPR000261; Eps15_homology.
 DR InterPro; IPR003903; UIM.
 DR Pfam; PF00036; ehand; 4.
 DR SMART; SM00054; Efh; 4.
 DR SMART; SM00027; EH; 3.
 DR SMART; SM00726; UIM; 3.
 DR PROSITE; PS00018; EF_HAND; 2.
 DR PROSITE; PS50031; EH; 3.
 DR PROSITE; PS50330; UIM; 2.
 KW Phosphorylation; Calcium-binding; Repeat; 3D-structure.
 FT DOMAIN 15 104 EH 1.
 FT DOMAIN 1 128 216 EH 2.
 FT CA_BIND 173 184 EF_HAND 1 (POTENTIAL).
 FT DOMAIN 224 314 EH 3.
 FT CA_BIND 236 247 EF_HAND 2 (POTENTIAL).
 FT DOMAIN 599 829 15 x 3 AA REPEATS OF D-P-F.
 FT REPEAT 599 601 1.
 FT REPEAT 623 625 2.
 FT REPEAT 629 631 3.
 FT REPEAT 634 636 4.
 FT REPEAT 640 642 5.

FT REPEAT 645 647 6.
 FT REPEAT 651 653 7.
 FT REPEAT 665 667 8.
 FT REPEAT 673 675 9.
 FT REPEAT 693 695 10.
 FT REPEAT 711 713 11.
 FT REPEAT 806 808 12.
 FT REPEAT 827 829 13.
 FT DOMAIN 852 871 UIM 1.
 FT DOMAIN 878 897 UIM 2.
 FT MOD_RES 850 850 PHOSPHORYLATION (BY EGFR).
 FT MUTAGEN 850 850 Y->F: INEFFICIENT EGFR INTERNALIZATION.
 SQ SEQUENCE 897 AA; 98470 MW; 08A0C0D423F873C2 CRC64;
 Query Match 7.8%; Score 486; DB 1; Length 897;
 Best Local Similarity 21.3%; Pred. No. 4.9e-12;
 Matches 205; Conservative 138; Mismatches 298; Indels 322; Gaps 31;
 15 WAITERAKHDQOFLKPIAGFITDQANFFQSGLPQVLAQIWMALADNNDGRMD 74
 122 MAVSEDKAKYDAIFDLSFVDFGLSDGKYKPVLNKLPVEILGRWELSDIDHDKLD 181
 75 QVEFSIAMKLIKLTQGYQLPSTLPVWKQOQVAISSAPAFGIGIASMPPLTAVAPVM 134
 182 RDEFAVAMFLVYCALF-----KEPVPM 203
 135 GSIVGKMSPLVSVPPAAPPVPLANGAPVIOPLPAFPAATWPKSSFSRSGPSQL 194
 204 -----SLPPALVP-----SKR 215
 195 NTKLQKQSPVVASAPAAEAWVPPQSRLKYROLFNHSDKTMGHTGPQARTLMQSL 254
 216 KT-----WVSPAKKATYDEFLTKTDMDGVYGLVRETFRTGL 257
 255 PQQASIAWNLSDIDQDKLTAEFFILAMLDIVM-SGQPLPVPVPEYIPSGFRVRS 313
 258 PSALLAHWSLCTDKGCKSLKQDFALFHLINQKLIKIDIPHSHPMPSPS---DRS 314
 314 GSGMSVSSSSVYQRLFEESSEDEQPEKLPTEPEDKKRENERERSVLEKROALLE 373
 315 SLQNTIGSSPV-----ADFAIK-----ELDTLNNEIVD 344
 374 QQRKEORLQLEAEERKERERQOEAKKQLEKQLEKQLEKQLEKQLEKQLEKQLEK 433
 345 LQREK-----NNVEDLKE-----DYKQKTSVQ---DLQVQRE 380
 434 EAARLEERQLEWERNRRQELNQRNKGQGTIVLAKRKTLFELALNDKKHQLQEG 493
 381 SINLQLOAQKQ-----QVQELLGE-----LDEQKQLE 410
 494 KLQDIRCLATQROEISTNKSRELRLAETHTLQOQLOESQOMGLRILPEKQILSDQK 553
 411 QLOVRKKCAEALISLK-----AETTSQSSQISYEELK-----ARELSR 456
 554 VQNSLHRDSLTLTKRLAEKELARQOLREQLDEVERE-----TRSKLOEIVFNNOKE 608
 457 LQQTAAQ-----LEESVESKQLEPLQCHLQSQEISSMQRLEMKQLFTNNQ--- 507
 609 LREIHSKQLOLQK-----RSLAARLQKQEKRSLELEKQEDAQRRVQ 653
 508 -----SNWSSPQGVLVNGATDYCSLTSSETANFHEAGQNNLSEPHQSSVRS 562
 654 ER-----DKQMLEHVQOEQPPRPKPHEDRLKRDYSVKKA-----ERAPENQDK 702
 563 PEIAPSDVTSEAVTVAQNEKVTPL--FDQDKHSKEEDPENNVSSTLTDVAADTNIDFF 620
 703 QSRLE-----HPHOEPAKLATQAPWSTTEGPTLISAOESVKKVYV 743
 621 QSDPFVSGDPKDPFGKIDPFGDPPFGKSPFASDQFPKQTSIDPFTSTSDPFSASN 680
 744 RALYPFESRSHDEITIGDVIWVDESQGTBP--GWLQGLKGTGWFPPANYAEKIDENE 801
 681 SSNTSVETWKNDPFAQGTIVAAASDSATDPFASVFGNESFG--DGFADFTLSKVNND 739

QY 802 VETP-----AKPVDLTSA-----PA--PKLALRENPAPLPTTSSEPTTPNN 842
 DB 740 ANNPRTSSSTSSVTAKPKMLBETASSEVPPALPKVGTTPRPCCPP-----PGKRPIN 794
 QY 843 WAFSST-----WSSSNKEPRTDNDMTWAQPSLTVPASAGQLRQSAFTPATATGSS 895
 DB 795 KLDSSDPLKLNDFQFQFPENDSPKENDPDMFC-----DPFTSSTTTNKE 838
 QY 896 PSP 898
 DB 839 ADP 841

RESULT 9
 PAN1_YEAST STANDARD; PRT; 1480 AA.
 AC P32521;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE PAN1 protein.
 GN PAN1 OR MIP3 OR MDP3 OR YIR006C OR YIB6C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OC NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 320-344; 352-375 AND 899-906.
 RX MEDLINE=92405166; PubMed=1339314;
 RA Sachs A.B., Dardorff J.A.;
 RT "Translation initiation requires the PAB-dependent poly(A)
 RT ribonuclease in yeast";
 RL Cell 70:961-973 (1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C;
 RX MEDLINE=95282515; PubMed=7762303;
 RA Voss H., Timmes J., Teodora C., Valencia A., Sengen C., Wiemann S.,
 RA Schwager C., Zimmermann J., Sander C., Ansoyge W.;
 RT "Nucleotide sequence and analysis of the centromeric region of yeast
 RT chromosome IX.";
 RL Yeast 11:61-76(1995).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C / AB972;
 RX PubMed=9169870;
 RA Churcher C.M., Bowman S., Badcock K., Bankier A., Brown D.,
 RA Chillingworth T., Connor R., Devlin K., Gentles S., Hamlin N.,
 RA Harris D.E., Hornell T., Hunt S., Jagels K., Jones M., Lye G.,
 RA Moulis S., Odell C., Pearson D., Rajandream M.A., Rice P., Rowley N.,
 RA Skelton J., Smith V., Walsh S., Whitehead S., Barrett B.G.;
 RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome IX";
 RL Nature 387:84-87(1997).
 CC -!- FUNCTION: NOT KNOWN.
 CC -!- PTM: THE N-TERMINUS IS BLOCKED.
 CC -!- SIMILARITY: Contains 2 EH domains.
 CC -!- CAUTION: WAS ORIGINALLY THOUGHT TO BE A SUBUNIT OF PAB-DEPENDENT
 CC POLY(A)-SPECIFIC RIBONUCLEASE.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; Z38062; CAAB6208.1; -
 CC EMBL; X79743; CAB38097.1; -
 CC EMBL; M90688; AAA34841.1; -
 CC PIR; S48440; S48440.
 CC SGI; S0001445; PAN1.

DR GO; GO:0005857; C:actin cortical patch (sensu Saccharomyces); IDA.
 DR GO; GO:0005886; C:plasma membrane; IDA.
 DR GO; GO:0008093; F:cytoskeletal adaptor activity; IPI.
 DR GO; GO:0007120; P:axial budding; IMP.
 DR GO; GO:0016288; P:cytokinesis; IMP.
 DR GO; GO:0006897; P:endocytosis; IMP.
 DR GO; GO:0007121; P:polar budding; IMP.
 DR InterPro: IPRO00261; EPI5_homology.
 DR Pfam: PF00036; ehband; 3.
 DR SMART; SM00054; EFh; 3.
 DR SMART; SM00027; EH; 2.
 DR PROSITE; PS50031; EH; 2.
 KW Repeat.
 FT DOMAIN 142 575 15 X 12 AA APPROXIMATE REPEATS.
 FT REPEAT 142 153 1-1.
 FT REPEAT 142 153 1-1.
 FT REPEAT 164 175 1-2.
 FT REPEAT 188 199 1-3.
 FT REPEAT 215 226 1-4.
 FT REPEAT 235 246 1-5.
 FT DOMAIN 270 359 EH 1.
 FT REPEAT 332 403 1-6.
 FT REPEAT 392 433 1-7.
 FT REPEAT 409 420 1-8.
 FT REPEAT 422 433 1-8.
 FT REPEAT 446 457 1-9.
 FT REPEAT 467 478 1-10.
 FT REPEAT 488 509 1-11.
 FT REPEAT 507 518 1-12.
 FT REPEAT 538 549 1-13.
 FT REPEAT 545 556 1-14.
 FT REPEAT 564 575 1-15.
 FT DOMAIN 328 680 2 X 23 AA APPROXIMATE REPEATS.
 FT REPEAT 328 350 2-1.
 FT DOMAIN 600 689 EH 2.
 FT REPEAT 600 689 2-2.
 FT DOMAIN 1084 1125 7 X 6 AA TANDEM REPEATS.
 FT REPEAT 1084 1089 3-1.
 FT REPEAT 1090 1095 3-2.
 FT REPEAT 1096 1101 3-3.
 FT REPEAT 1102 1107 3-4.
 FT REPEAT 1114 1113 3-5.
 FT REPEAT 1120 1125 3-6.
 FT DOMAIN 1315 1377 8 X 6 AA APPROXIMATE REPEATS.
 FT REPEAT 1315 1320 4-1.
 FT REPEAT 1321 1326 4-2.
 FT REPEAT 1327 1332 4-3.
 FT REPEAT 1340 1345 4-4.
 FT REPEAT 1345 1350 4-5.
 FT REPEAT 1355 1360 4-6.
 FT REPEAT 1361 1366 4-7.
 FT REPEAT 1372 1377 4-8.
 FT DOMAIN 13 22 POLY-GLN.
 FT DOMAIN 29 34 POLY-GLN.
 FT DOMAIN 98 106 POLY-GLN.
 FT DOMAIN 140 1406 POLY-PRO.
 FT DOMAIN 1452 1455 POLY-GLU.
 FT DOMAIN 1474 1480 POLY-PRO.
 FT CONFLICT 265 235
 FT CONFLICT 266 273
 FT CONFLICT 474 487
 FT CONFLICT 653 657
 FT CONFLICT 1291 1291
 FT CONFLICT 1396 1480

SO SEQUENCE 1480 AA; 160267 MW; F3518435FF59553 CRC64;
 Query Match 6.5%; Score 410.5; DB 1; Length 1480;
 Best Local Similarity 20.5%; Pred. No. 6.1e-09;
 Matches 261; Conservative 174; Mismatches 444; Indels 395; Gaps 47;

```

OY 17 ITVERAKHQDQFLSLKPI-AGFITGQDARNFPOFSGSLPQVLAQIWMALAMNNDGRDQ 75
Db 266 ITADQDAKFETLFSIVTNSNTVSGANCKILMRSGLPQSLARIMTLCDTSTAGELLF 325
OY 76 VEFISAMKLIKLIKQYQPLPSTLPPVMKQ----- 104
Db 326 PEFALAMLINDVLOGDPIPYELDSKTNEVSSFIDAINLSIANQDSSANDAPKTPDEF 385
OY 105 -----QPVASAPAFGI-----GGLASM----- 123
Db 386 ITAGVONLQOPQPTGYMPQTSFGIPLQSQITGGVASALNPOSTGFMATPTENSMNTGTP 445
OY 124 ---PPLTAVAVPM-----GSIP-VVGMSPPLVSSVPPAAV-----PPLAN 160
Db 446 GLNPQITGAPASMQPNITGNALQPTTGMMPQTTGMMPQTTGMMPQTSFGVNLGFLTG 505
OY 161 GAPPVIGP-----LPFAHAPAAVWPXSSFSRSGPSQL----- 194
Db 506 GA---LQSQYTGYSVMPQSGPASM---PNLSFMOQLQSLTGLQPTGFLPPSNFS 560
OY 195 -----NTKQKASFDVASAPPAEAMVPPSSRLKXROLNSHDKTMSGLT 241
Db 561 ATMPLTACTGFGNNEIYTKSNFNNLIDNSSQDKISTEEKSLFKIETEDTQNGKLLD 620
OY 242 GPOARTILMOGSLPOAQLASIMNLSIDIDQDGLTAEFFILAMHLIDVAMSGPLPPVLP 301
Db 621 SPTAVEIRKSGILRADLEQIMNLCIDINNTQNLNQBALMHLVYGLNKGKPIPNVLPS 680
OY 302 EYIP-----DSF-----RRVRSQ- 314
Db 681 SLIPSTKLNLNKLKTEPTTTEKESFGKIDALSYKNDDVLPVYRRRKVYSAKN 740
OY 315 ---SGMSITSSSVUDQRLPEBSSSDEQPEKLPVTEDEK---KRNFERGSVELEK 367
Db 741 EEOSSFSPPSAKSVNH-----SSSTLTQTDISVKTVEKTKAKYAGFR---EIMLK 791
OY 368 ROALLEQORKEOR-----LAQERABOEKKEBEOEAKROLE 407
Db 792 NIASLENITKINISNENCYDSSIPSDLTSRPAILAKPMLFNEISTIDNETAKIDLY 851
OY 408 LEKQLEKQRELERQEEERREKREIEREAKRELEROROLEWERNROELNQRNKEOCT 467
Db 852 RKK---NPSIISGSPNGEITENDRKAKKASRLALR-----ARMALTGKSTESBDS 899
OY 468 VILKARKRTILEFE-LEALNDKQHOLEGLQDIDRCLATQROEIBESTNSRELIAETIHL 526
Db 900 LSMEDEQOSAEIKRIQOENGNQOET---IKDIRSSISISASLKSTMTGSMN---ISN- 951
OY 527 QOQLOESQOMLGRLLPEKQILSDQKQVQONSIAHDS----- 563
Db 952 -QEFERMEFGIGLEGVAREFL-DLJKSNSKSVTSSSPFVBSSTTPVDDSSSPSYGOF 1009
OY 564 LITLRALAEKELARQOLREOLDEVERETRSKLQSIDVFN-NQKELREIHSKOOL--- 618
Db 1010 KTAEBRAAYLVKEQAKKMKELAKFDKRNRYNTQSSRSISSBNSREOQIAGSSNLVBP 1069
OY 619 -----QKRSLEAALKQKQERKSLLEKQKE----- 646
Db 1070 RATPQEEKYEVAPQTPVOSTQVQPTQVQPTQVQPTQVQPTQVQPTQVQVNY 1129
OY 647 DAQRVQERDKQWLEHVQOEQPRPKPHEEDRLR-----BDSVRKKEAE- 693
Db 1130 NAKQSDDEDEDEDEKRLQEBLKLKLLKKADKELKALAKQIIDAQNESBEEBTNGKD 1189
OY 694 -----RAKP-----EMODKOSRLFPHOEP- -KLATQAPWSTTEKGLTISA 734
Db 1190 NFGHVAVPQAPFAPVAPSAFQNSNTAPRSVAAVTPAAGKNSGTLPSTTMGHNVYFQDA 1249
OY 735 QESVAVVYVYKALYPRESHD-EITIQPDIVMVBESQTEGFWGELKKGKTGMFPANY 793
Db 1250 SAS-----STSTPDARAEMQRIQOR---LDEDE---DDGMSDEDESNNR---VAV 1292
OY 794 AEKIPENEVPTP---AKVTDLTAPAPKALARETAP- -LPVTSSEPTTPNNWADFS 847

```

```

Db 1293 DNKYEAKIGHPDRAPV---TAAPLPSV---TVPVAVPQANTSNKSSPIPIA 1345
OY 848 STWPSSNE-----KEPTDMWDTAAOPSLTVSAGLRORS-AFTPATATGSSPSVLG 901
Db 1346 PIPPSVTOEBPPVPLAPLPAVDGQEBPPIPSAPAIATVAKSGSSTPALAGVLPP- 1402
OY 902 QGEKVEGLQAOALYPMKAKDNHLENFK-----SDVITVLEQQDMWFGCEVQOK 951
Db 1403 --PPLPQOASTSEPTIAHVDNNGAEKNGAYGSDSDVVLSPES---VCTDEEB 1455
OY 952 GWPEKSYVKLISGP 965
Db 1456 GAQPVSTAGIPISIP 1469

RESULT 10
YBET YEAST STANDARD; PRT; 1381 AA.
ID YBET YEAST
AC P34216;
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Hypothetical 150.8 kDa protein in SEC17-QCRI intergenic region.
GN YBL047C OR YBL0520 OR YBL0501.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE OF 1-961 FROM N.A.
RC STRAIN=5288C;
RX MEDLINE=95176707; PubMed=7871888;
RA de Wergifosse P., Jacques B., Jomiaux J.-L., Purnelle B., Skala J.,
RA Goffeau A.;
RT "The sequence of a 22.4 kb DNA fragment from the left arm of yeast
RT chromosome II reveals homologues to bacterial proline synthetase and
RT murine alpha-adaptin, as well as a new permease and a DNA-binding
RT protein."
RL Yeast 10:1489-1496(1994).
RN [2]
RP SEQUENCE OF 579-1381 FROM N.A.
RC STRAIN=5288C;
RX MEDLINE=94205266; PubMed=8154187;
RA Scherens B., el Bakhouy M., Vierendeels F., Dubois B., Messenguy F.;
RT "Sequencing and functional analysis of a 32,560 bp segment on the
RT left arm of yeast chromosome II. Identification of 26 open reading
RT frames, including the XIPI and SEC17 genes."
RL Yeast 9:1355-1371(1993).
CC -1- SIMILARITY: BELONGS TO THE VDP/USO1/YBL047C FAMILY.
CC -1- SIMILARITY: Contains 3 EH domains.
CC -1- SIMILARITY: Contains 1 UBA domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.1db-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: Z35808; CA84867.1; -
DR EMBL: X78214; CA85048.1; -
DR EMBL: Z23261; CA880797.1; -
DR PIR: S45781; S45781.
DR SCG: S0000143; EDEL.
DR GO: GO:0005857; C:actin cortical patch (sensu Saccharomyces); IDA.
DR GO: GO:0005935; C:bud neck; IDA.
DR GO: GO:0005934; C:bud tip; IDA.
DR GO: GO:0006837; P:endocytosis; IMP.
DR InterPro: IPR000261; EPS15_homology.
DR InterPro: IPR000449; UBA_domain.
DR Pfam: PF00036; ehand; 2.

```

DR Pfam; PF00627; UBA; 1.
 DR SMART; SM00054; Efh; 2.
 DR SMART; SM00027; Efh; 3.
 DR SMART; SM00165; UBA; 1.
 DR PROSITE; PS50031; Efh; 3.
 DR PROSITE; PS50030; UBA; 1.
 DR Hypothetical protein; Repeat.
 FT DOMAIN 14 113 EH 1.
 FT DOMAIN 135 227 EH 2.
 FT DOMAIN 277 366 EH 3.
 FT DOMAIN 1338 1380 UBA.
 SQ SEQUENCE 1381 AA; 150783 MM; 626FD261DCBA7D99 CRC64;

Query Match 6.4%; Score 403; DB 1; Length 1381;
 Best Local Similarity 20.0%; Pred. No. 1.1e-08;
 Matches 266; Conservative 199; Mismatches 486; Indels 458; Gaps 57;

3 QEPTEP-GGS-----LDVMTVTERAKHDQFLSLKPIAGFITGDARNFQSGHP 54
 109 QNPAPMQSGSATGNTNTNDI PALSSNDIAKFSQLPRTAKGAQTAVGDKAKOIFLKARLP 168
 55 QGVLAQIALAMNDGRMDQVEFSIAMLKILKLGQYQPLSTLPVMKQAPVALISSADA 114
 169 NQTLGEMALCDRDASGVLDKSEFFIMAWYLQICMSHSPSMTTPPAVLPTQ----- 219
 115 FGIGGIASMPPLTAAVAVPMGSIPIVVGMSPIVSSVPPAAVPLANGAPPVLIQPLPAFAH 174
 220 -----LMDSI-----RLEPVVNVQNPRTTTPISANST 245
 175 PAATWPKSSPSRSGSQLNTKLQKASQFDVASAPPAEWAVPQSSRLKRYQLFNSHDK 234
 246 GVSLSLRHSTIRLSLGAFA--ASDMSLFEKKQDPAIFPSLDK 290
 235 TMSGHITGQATITLMQSSLPOALASINLSDIDODGKLTAEFTLANHLIDVANSQGP 294
 291 QHAGSISSAVLVFPFLSSRLNQTETATIDLADIHNNAEFTKLEFALIAFLIOKKNAGVE 350
 295 LPPVLPE-----YIPSPFRVRSGSGMSVSSSS----- 324
 351 LDDVLPNELSGPALGLYPNPLPQQQSAPOIALIPRASKPSLQDMPHQVSAVANTOPT 410
 325 VDQRLPE-----EPS-----SEDEQOP 341
 411 VPQVLPQNSNGSLNDLALNPSFSPPTKATQVYVQNTNNSFSYDNNNGCATILQQQOP 470
 342 EKKLPVTEDEKRENFERSGVELERKQALLERKQERL----- 382
 471 QQPPPLTHSSSGLKFTPTS---NFGQSIILKEEPERQULRESSDPTFSAOPPPVKHAS 526
 383 ----- 382
 527 SPVKRTASTTLPOVPNFVSFMPAGATSAATGAAGAAVGAALGASAFSSSSNNAFKN 586
 383 -----AQLERAQERKERERQEOEAKROLLEKOLEKORLELEROERERREKIR 432
 587 QDLFADGESAQLSNATTEMANLSNVNLSKQASTNDKKSRAITDELKRVTEMKNSIOI 646
 433 REAAKRELERQOLEWERNRQELLNQRNKEQEGTVILAKRKTIIELEALNDKKQOLE 492
 647 KLANLRSTIDQVVKQTEQLEAQL--QVKNEN-----TLAQQLAVEANYHAHE 694
 493 GKLQDIRCLATQROEISTNSRELRIMEI-----THLOOQLOESQMGRLIPEKQILS 548
 695 SKLNE---LTTDLQSSQTKNAELKEQITNLNLSMTASLOSQLEKQO----- 737
 549 DQLKVOQNSLHRDSLTLTKRALAEKELARQOLEQLEDEVERETSKLOEIDVFNQOLKE 608
 738 -QVKQ-----ERSWVDVNSKQLELNQVYVNLQKIDLG-----EKISVYLTKQEH 783
 609 LREIHSKQOLQKRSLEAARLKQKEOE--RKSLET--EKQKEDAQRVQERDKQMLEHV 663
 784 LND-----YQKTVBERQHAQOLQAKYQDLSNKKQDITLDRKQOLEERNRQLEQEBENIYHQV 837

QY 664 QOEOBPRKPRHEERLKRDESVRKKAERAK-----PEMOKOSRLFHHPQEP-- 713
 DB 838 SKLQF-----MFDLSQKASFEKADQELKERNIEYANNVRELSEKQNL-AMGQLPED 890
 QY 714 -----AKLATQAPWST---TEKGPL---TIS-----AOESVYVYVYRALYPRESSHDEI 757
 DB 891 AKDIIAKSASNTDITTKKATSRGNVHEDTVSAFVETTVENSUNLVNRRVADDEKERTES 950
 QY 758 TIQPGDI-VMVDESQTEGPGMLGSLKGGTKGTFPANYAEKIP-----NEVPPT- AKV 809
 DB 951 DVFDRDVPFLGQSQSENANTNNGTQSGNETANP-NLTETLSDRFQDLENGVIPSQGL 1009
 QY 810 TDLTAPAPKALRETPAPLPVTSSSEPTTPANNWADFSTWSSSNEKETONMDTWAQ 869
 DB 1010 TSSVANNAPO-SVRD-DVELPETERDIT-WNTANRDT-----GNLSHIPEEWETPAT 1062
 QY 870 PSALTVPASQQLQRSAFTPATATGSSPSVLG--QGEKEGLOAQ---ALYPMRAKQD- 923
 DB 1063 ASTDVL-----NETTEVIDGSTTKRANSNEDGESVSIQESPKLSAQ-KAKTIVE 1114
 QY 924 -----HLNFKSDVITVLEQQDMMWFGVQCKMPKSYVKLISGP----- 965
 DB 1115 EPPPIQELHIDSDSSDDDE---FEDTRE---IPSAIVTKLQTPYNAOPTSLEIH 1166
 QY 966 ---VRKSTSIDTPESPASLKRVAAPAKAIP--GEEFIMTYEESQGDLPFGQGD 1020
 DB 1167 TEQVTKYPAPGTPSPHNBGNSKKASTNLTLPKXDEPDEFAGL-----EQAAVEEDNG- 1219
 QY 1021 VIVVTKQGDMMWTGTVGKSGVFPANNVRLKQSE-----GSGTAGKTSLGKPEI 1071
 DB 1220 -----ABSESEFENVANANGSMEQFETIDHOLDDELQNNATGTLTSSNPTI-PKQV 1272
 QY 1072 AQVIASVYATGPEQOLTLAPGOLILRKKNPGSGWEGELQARCKRQIGFPPANYVKLSP 1131
 DB 1273 QO-----QSTSDPAQV-----SNDEWD-EIIFA-----GF----- 1295
 QY 1133 GTSKITPEEL--PKTAVPAVCQVIGMYDTQOND--ELARSKQIIN 1176
 DB 1296 GNSKAEPTKVATPTSPQOP-----IPLKNDPIVDASLSKQPTVN 1334

RESULT 11
 TRHY_RABIT
 ID TRHY_RABIT STRAND: PRT; 1407 AA.
 AC P37709;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Trichonyalin.
 GN THH.
 OS Oryctolagus cuniculus (Rabbit).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
 OC NCBI_Taxid=9986;
 RP [1]
 RA SEQUENCE FROM N.A.
 RA Fletz M.J., Rogers G.E.;
 RA Submitted (DEC-1992) to the EMBL/Genbank/DBJ databases.
 CC -FUNCTION: INTERMEDIATE FILAMENT-ASSOCIATED PROTEIN THAT ASSOCIATES
 CC -IN REGULAR ARRAYS WITH KERATIN INTERMEDIATE FILAMENTS (KIF) OF THE
 CC INNER ROOT SHEATH CELLS OF THE HAIR FOLLICLE AND THE GRANULAR
 CC LAYER OF THE EPIDERMIS. IT LATER BECOMES CROSS-LINKED TO KIF BY
 CC ISOPEPTIDE BONDS. IT MAY SERVE AS SCAPFOLD PROTEIN, TOGETHER
 CC WITH INVOLUCRIN, IN THE ORGANIZATION OF THE CELL ENVELOPE OR EVEN
 CC ANCHOR THE CELL ENVELOPE TO THE KIF NETWORK. IT MAY BE INVOLVED IN
 CC ITS OWN CALCIUM-DEPENDENT POSTSYNTHETIC PROCESSING DURING TERMINAL
 CC DIFFERENTIATION.
 CC -SUBUNIT: Homodimer (Probable).
 CC -TISSUE SPECIFICITY: FOUND IN THE HAND KERATINIZING TISSUES SUCH AS
 CC THE INNER ROOT SHEATH (IRS) OF HAIR FOLLICLES AND MEDULLA, AND IN
 CC THE FILIFORM PAPILLAE OF DORSAL TONGUE EPITHELIUM (PROBABLE).
 CC -DEVELOPMENTAL STAGE: EXPRESSED DURING LATE DIFFERENTIATION OF
 CC THE EPIDERMIS.


```

CC -1- DOMAIN: CONSISTS OF NINE DOMAINS. DOMAIN 1 CONTAINS TWO EF-HAND
CC CALCIUM-BINDING DOMAINS. DOMAINS 2-4, 6, AND 8 ARE ALMOST
CC ENTIRELY ALPHA-HELICAL, CONFIGURED AS A SERIES OF PEPTIDE REPEATS
CC OF VARYING REGULARITY, AND ARE THOUGHT TO FORM A SINGLE-STRAINED
CC ALPHA-HELICAL ROD STABILIZED BY IONIC INTERACTIONS. DOMAIN 6 IS
CC THE MOST REGULAR AND MAY BIND KIP DIRECTLY BY IONIC INTERACTIONS.
CC DOMAINS 5 AND 7 ARE LESS WELL ORGANIZED AND MAY INDUCE FOLDS IN
CC THE MOLECULE. DOMAIN 9 CONTAINS THE C-TERMINUS, CONSERVED AMONG
CC DIFFERENT SPECIES.
CC -1- PTH: SUBSTRATE OF TRANSGUTAMINASE. SOME 200 ARGININES ARE
CC PROBABLY CONVERTED TO CITRULLINES BY PEPTIDYLARGININE DEIMIDASE.
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE S-100
CC FAMILY.
CC -1- SIMILARITY: Contains 2 EF-hand calcium-binding domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; 219092; CAA79519.1; -
CC PIR; S28589; S28589.
CC HSSP; P02633; 41CB.
CC InterPro; IPR001751; CAP_S100.
CC InterPro; IPR002048; EF-hand.
CC Pfam; PF00036; ehand; 1.
CC Pfam; PF01023; S_100; 1.
CC ProDom; PD003407; CAP_S100; 1.
CC PROSITE; PS00018; EF_HAND; 1.
CC PROSITE; PS00303; S100_CAP; 1.
CC KeraTat1zation; Calcium-binding; Repeat; Citrullination.
CC FT DOMAIN 1 91 S-100 LIKE.
CC CA BIND 22 33 EF-HAND 1 (LOW AFFINITY) (POTENTIAL).
CC FT CA BIND 62 73 EF-HAND 2 (HIGH AFFINITY) (POTENTIAL).
CC SQ SEQUENCE 1407 AA; 183781 MW; AEL7DA159P1287F CRG4;

Query Match 6.2%; Score 391.5; DB 1; Length 1407;
Best Local Similarity 29.7%; Pred. No. 3.1e-08;
Matches 132; Conservative 97; Mismatches 140; Indels 75; Gaps 18;

QY 326 DQRLPESSSEDEQOEKKLPVTFEDKRENERGVSLEKRRQLLEEQ----- 375
DB 198 EEFIEEQLRRREQLRELEEEQQRREERREHERALQEEBQLKQRWRREPRRQ 257
QY 376 -----RKEQRIALQLEAEQERKERERQOEAKQLELEKQLEKQRELERQREER 426
DB 258 QLRRLERIRERREQLREERREQLREERREQLREERREQLREERIRERREQLRE 317
QY 427 RKE-----TERRE-AAKREL---ERQOLEWERNRROELLNQRKEQ---EGTVVLR 473
DB 318 RREQLREERREQLREERREQLREERREQLREERREQLREERREQLREERREQL 376
QY 474 RKTLEFELALNDKKGQLEKQDRCRLATQROEIESNRSRELRIMEITLQOQL-- 531
DB 377 QRLSESEGAROSKYSRPRROESLSLQDOERRO-----ROKRELEBQARRQOMQAE 432
QY 532 -ESQMLRL-----IPKQLSDQLKQVQ-----NSLHRDSTLTLRALAEKLARQ 580
DB 433 ESEERRRRLSARPLRLRQLAERROEOFRFEERRRRRRRELQFLREBEOQLRE 492
QY 581 LREQLDE-----VERETSQKQEI--DVFNQQLKE--LREIHS-----KQQLQKRS 623
DB 493 RAQQLQEDSEFQEDERRRRRQOEGRPGQTWQQLQEEAKRRHTTLYAKPGQOEQLRE 552
QY 624 LEAALTKQEQKRSLELEK--QKEDAQRRVQERKQW--LEHVQOEOPRRKREHEDR 679
DB 553 LQRERRRO-ERRERREBEKQLQREDEKRRRQEREROYLELEBEOQLRDKRLEBQ 611
QY 680 L---KREDSVRKKEAEERAKPEMQ 700

```

```

DB 612 LQREERERLRQERERKREERQ 635

RESULT 12
ID TRHY HUMAN STANDARD; PRT; 1898 AA.
AC 007283;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Trichohyalin.
GN THH OR TRHY OR THL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93280194; PubMed=7685034;
RA Lee S.-C., Kim I.-G., Marekov L.N., O'Keefe E.J., Parry D.A.D.,
RA Steinert P.M.;
RT "The structure of human trichohyalin. Potential multiple roles as a
RT functional EF-hand-like calcium-binding protein, a cornified cell
RT envelope precursor, and an intermediate filament-associated (cross-
RT linking) protein."
RL J. Biol. Chem. 268:12164-12176(1993).
RN [2]
RP SEQUENCE OF 1731-1898 FROM N.A., AND CHARACTERIZATION.
RX MEDLINE=93315897; PubMed=7686953;
RA O'Keefe E.J., Hamilton E.H., Lee S.-C., Steinert P.M.;
RA "Trichohyalin: a structural protein of hair, tongue, nail, and
RT epidermis."
RL J. Invest. Dermatol. 101:655-715(1993).
CC -1- FUNCTION: INTERMEDIATE FILAMENT-ASSOCIATED PROTEIN THAT ASSOCIATES
CC IN REGULAR ARRAYS WITH KERATIN INTERMEDIATE FILAMENTS (KIF) OF THE
CC INNER ROOT SHEATH CELLS OF THE HAIR FOLLICLE AND THE GRANULAR
CC LAYER OF THE EPIDERMIS. IT LATER BECOMES CROSS-LINKED TO KIF BY
CC ISODIPEPTIDE BONDS. IT MAY SERVE AS SCAFFOLD PROTEIN, TOGETHER
CC WITH INVOLUCRIN, IN THE ORGANIZATION OF THE CELL ENVELOPE OR EVEN
CC ANCHOR THE CELL ENVELOPE TO THE KIF NETWORK. IT MAY BE INVOLVED IN
CC ITS OWN CALCIUM-DEPENDENT POSTSYNTHETIC PROCESSING DURING TERMINAL
CC DIFFERENTIATION.
CC -1- SUBUNIT: Monomer (Probable).
CC -1- TISSUE SPECIFICITY: FOUND IN THE HARD KERATINIZING TISSUES SUCH AS
CC THE INNER ROOT SHEATH (IRS) OF HAIR FOLLICLES AND MEDULLA, AND IN
CC THE FLITIFORM PAPILLAE OF DORSAL TONGUE EPITHELIUM (PROBABLY).
CC -1- DEVELOPMENTAL STAGE: EXPRESSED DURING LATE DIFFERENTIATION OF
CC THE EPIDERMIS.
CC -1- DOMAIN: CONSISTS OF NINE DOMAINS. DOMAIN 1 CONTAINS TWO EF-HAND
CC CALCIUM-BINDING DOMAINS. DOMAINS 2-4, 6, AND 8 ARE ALMOST
CC ENTIRELY ALPHA-HELICAL, CONFIGURED AS A SERIES OF PEPTIDE REPEATS
CC OF VARYING REGULARITY, AND ARE THOUGHT TO FORM A SINGLE-STRAINED
CC ALPHA-HELICAL ROD STABILIZED BY IONIC INTERACTIONS. DOMAIN 6 IS
CC THE MOST REGULAR AND MAY BIND KIP DIRECTLY BY IONIC INTERACTIONS.
CC DOMAINS 5 AND 7 ARE LESS WELL ORGANIZED AND MAY INDUCE FOLDS IN
CC THE MOLECULE. DOMAIN 9 CONTAINS THE C-TERMINUS, CONSERVED AMONG
CC DIFFERENT SPECIES.
CC -1- PTH: SUBSTRATE OF TRANSGUTAMINASE. SOME 200 ARGININES ARE
CC PROBABLY CONVERTED TO CITRULLINES BY PEPTIDYLARGININE DEIMIDASE.
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE S-100
CC FAMILY.
CC -1- SIMILARITY: Contains 2 EF-hand calcium-binding domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; L09190; AAA65582.1; -
CC PIR; A45973; A45973.

```


VARYING REGULARITY, AND ARE THOUGHT TO FORM A SINGLE-STRANDED ALPHA-HELICAL ROD STABILIZED BY IONIC INTERACTIONS. DOMAIN 6 IS THE MOST REGULAR AND MAY BIND KIT DIRECTLY BY IONIC INTERACTIONS. DOMAINS 5 AND 7 ARE LESS WELL ORGANIZED AND MAY INDUCE FOLDS IN THE MOLECULE. DOMAIN 9 CONTAINS THE C-TERMINUS, CONSERVED AMONG DIFFERENT SPECIES.

-1- PM: SUBSTRATE OF TRANSGLUTAMINASE. SOME 200 ARGININES ARE PROBABLY CONVERTED TO CITRULLINES BY PEPTIDYLARGININE DEIMIDASE.

-1- SIMILARITY: IN THE N-TERMINAL SECTION, BELONGS TO THE S-100 FAMILY.

-1- SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/or_send_an_email_to_licensing@isb-sib.ch).

CC EMBL; Z18361; CA79165.1; -
 CC EMBL; X51695; CA35992.1; -
 CC PIR; A40691; A40691.
 CC HSSP; P02633; 11G5.
 CC InterPro: IPR001751; CAPB_S100.
 CC InterPro: IPR002048; EF-hand.
 CC Pfam; PF00036; ethand, 1.
 CC Pfam; PF01023; S_100; 1.
 CC ProDom; PD003407; CAPB_S100; 1.
 CC PROSITE; PS00018; EF_HAND; 1.
 CC PROSITE; PS00303; S100 CAPB; FALSE NEG.
 CC KEGG; K04003; Calcium-binding; Repeat; Citrullination;
 KW Alternative splicing.

FT DOMAIN 1 91 S-100 LIKE.
 FT CA_BIND 22 33 EF-HAND 1 (LOW AFFINITY) (POTENTIAL).
 FT CA_BIND 62 73 EF-HAND 2 (HIGH AFFINITY) (POTENTIAL).
 FT DOMAIN 413 832 14 X 28 AA APPROXIMATE TANDEM REPEATS.
 FT REPEAT 413 448 1-1.
 FT REPEAT 449 476 1-2.
 FT REPEAT 477 504 1-3.
 FT REPEAT 505 532 1-4.
 FT REPEAT 533 560 1-5.
 FT REPEAT 561 588 1-6.
 FT REPEAT 589 616 1-7.
 FT REPEAT 617 644 1-8.
 FT REPEAT 645 678 1-9.
 FT REPEAT 679 706 1-10.
 FT REPEAT 707 742 1-11.
 FT REPEAT 743 771 1-12.
 FT REPEAT 772 796 1-13.
 FT REPEAT 797 832 1-14.
 FT DOMAIN 938 1507 23 X 23 AA APPROXIMATE TANDEM REPEATS.
 FT REPEAT 938 961 2-1.
 FT REPEAT 962 985 2-2.
 FT REPEAT 986 1021 2-3.
 FT REPEAT 1022 1044 2-4.
 FT REPEAT 1045 1067 2-5.
 FT REPEAT 1068 1090 2-6.
 FT REPEAT 1091 1121 2-7.
 FT REPEAT 1122 1144 2-8.
 FT REPEAT 1145 1167 2-9.
 FT REPEAT 1168 1197 2-10.
 FT REPEAT 1198 1227 2-11.
 FT REPEAT 1228 1250 2-12.
 FT REPEAT 1251 1273 2-13.
 FT REPEAT 1274 1296 2-14.
 FT REPEAT 1297 1319 2-15.
 FT REPEAT 1320 1342 2-16.
 FT REPEAT 1343 1368 2-17.
 FT REPEAT 1369 1391 2-18.
 FT REPEAT 1392 1416 2-19.
 FT REPEAT 1417 1439 2-20.
 FT REPEAT 1440 1461 2-21.

FT REPEAT 1462 1484 2-22.
 FT REPEAT 1485 1507 2-23.
 FT VARSPLIC 1145 1197 Missing (in isoform Short).
 FT VARSPLIC 1251 1273 /FTID=VSP_000847.
 FT VARSPLIC 1251 1273 Missing (in isoform Short).
 FT VARSPLIC 1251 1273 /FTID=VSP_000848.
 FT CONFLICT 1399 1399 E -> G (IN REF. 2).
 FT SEQUENCE 1549 AA; 201173 MW; 872F89FF126554E CRC64;
 SQ

Query Match 6.0%; Score 375.5; DB 1; Length 1549;
 Best Local Similarity 27.6%; Pred. No. 1.4e-07;
 Matches 116; Conservative 101; Mismatches 114; Indels 89; Gaps 16;

QY 332 EPSSEDSQPEKXLPVTFE-----DKRENFGSVLEKRRQALLDQKQERLAQER 387
 DB 383 KPAQREYREBEQULRLAEKQLQKREKQREYREVLQRE--ERLQREBEQLOREER 440
 QY 388 AEQKREKREKQ-----DOAKRQLEKQ-----LEKQRELEKREBER 426
 DB 441 EKRRQREKQYLEKVELMEBEQLOREERERKQREKQYLEKVELMEBEQLOREER 500
 QY 427 RKEIERREBAKRELEKQREKQREKQREKQREKQREKQREKQREKQREKQREKQ 486
 DB 501 ROERREYLEKVELQREBEQLOREERERK-----RQRE-----ROYLE----- 538
 QY 487 KKHLEKGLDIDIRLATQREKQREKQREKQREKQREKQREKQREKQREKQREKQ 544
 DB 539 -KVELQREBEQ-----LQREKREKQREKQREKQREKQREKQREKQREKQREK 590
 QY 545 QILSDQLKQVQNSLRHDSLLTLKRALEKELAKQRLQREQLDEVERSTRKQLEIDVFN 604
 DB 591 QYL-EKVELQREBEQLOREKREKQREKQREKQREKQREKQREKQREKQREKQREK 641
 QY 605 QKLEKREHKSQQLQKQREKQREKQREKQREKQREKQREKQREKQREKQREKQ 650
 DB 642 QERREYLEKVELQREBEQLOREBEKQREKQREKQREKQREKQREKQREKQREK 701
 QY 651 RVQERDKQWL-EHVQREBEQRPKPEHDLKXEDSVKKEABER--AKPEMDQKQRL 706
 DB 702 RQERREYLEKVELQREBEQ-----RQREKQREKQREKQREKQREKQREKQREK 758

RESULT 14
 CALD_HUMAN STANDARD; PRT; 793 AA.
 AC Q05682; Q13978; Q13979; Q14741; Q14742;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 GN Caldesmon (CDM).
 OS CALD1 OR CDM OR CAP.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OX NCBI_TaxId=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 4).
 RC TISSUE=Lung fibroblast.
 RX MEDLINE=91358497; PubMed=1885618;
 RA Novy R.E., Lin J.L.-C., Lin J.J.-C.,
 RT "Characterization of cDNA clones encoding a human fibroblast
 RT caldesmon isoform and analysis of caldesmon expression in normal and
 RT transformed cells";
 RL J. Biol. Chem. 266:16917-16924(1991).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 4).
 RC TISSUE=Aorta;
 RX MEDLINE=92209999; PubMed=1555769;
 RA Humphrey M.B., Herrera-Sosa H., Gonzalez G., Lee R., Bryan J.,
 RT "Cloning of cDNAs encoding human caldesmons";
 RL Gene 112:197-204(1992).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORMS 2; 3 AND 5).


```

Oy 684 DSVRKEAEERAKPEMOD-----KQSRLEPHHQBPAKLATQAPWSTTEKGLTISAGES 737
Db 608 IERRAEAEAKRQKPEDEGLSDDKKPFKCTFKGSKLKEERAE-----LNKSVOKS 660
Oy 738 --VVVVVYRALT-PFESRSHDEITLQPDIVWDESGTGEPCWLGELKGTGFPPANYA 794
Db 661 SGVSTHQALVSKIDSLREQYTS-----ALBETKAKP----- 694
Oy 795 EKIPENVTPTAKPYTDL-----TSAPAP-----KLAL----- 822
Db 695 TKPAASDLPVPAEGFRNKSMWKEGKNVSPFAACTPKETAGLKVGVSRINEWLTXTYP 754
Oy 823 --RETPAPLPVTSSEPTTPNNMADFSSTWPSSEK 857
Db 755 DGNKSPAPKP-----SDLRPGDVSKRMLMEKQSVDK 786

RESULT 15
MAK4_HUMAN
ID MAK4_HUMAN STANDARD; PRT: 1239 AA.
AC O95819.075172; O9NST7;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Mitogen-activated protein kinase kinase kinase 4 (EC 2.7.1.37)
DE (MAPK/ERK kinase kinase kinase 4) (MEKKK 4)
DE (HRK/GCK-like kinase HGK) (NCK interacting kinase).
GN MAPK4 OR HGK OR NIK OR KIA0687.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Metazoa; Primates; Catarrhini; Homiidae; Homo.
OC NCBI_Taxid:9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 4 AND 5), FUNCTION, AND TISSUE
RP SPECIFICITY.
RC TISSUE=Macrophage;
RX MEDLINE=99107863; PubMed=9890973;
RA Yao Z., Zhou G., Wang X.S., Brown A., Diener K., Gan H., Tan T.-H.;
RT "A novel human STE20-related protein kinase, HGK, that specifically
RT activates the c-Jun N-terminal kinase signaling pathway.";
RL J. Biol. Chem. 274:2118-2125(1999).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Glycoblactoma;
RX PubMed=12612079;
RA Wright J.H., Wang X., Manning G., Lavere B.J., Le P., Zhu S.,
RA Khatri D., Flanagan P.M., Buckley S.D., Whyte D.B., Howlett A.R.,
RA Bischoff J.R., Lipson K.E., Jallat B.;
RT "The STE20 kinase HGK is broadly expressed in human tumor cells and
RT can modulate cellular transformation, invasion, and adhesion.";
RL Mol. Cell. Biol. 23:2068-2082(2003).
RN [3]
RP SEQUENCE OF 36-1239 FROM N.A. (ISOFORM 2).
RC TISSUE=Brain;
RX MEDLINE=98403880; PubMed=9734811;
RA Ishikawa K.-I., Nagase T., Suyama M., Miyajima N., Tanaka A.,
RA Kozumi H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. X.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL DNA Res. 5:169-176(1998).
RN [4]
RP SEQUENCE OF 262-1239 FROM N.A. (ISOFORM 3).
RC TISSUE=Testis;
RX Duecherhoef A., Lauber J., Mewes H.-W., Gaassenhuber J., Wiemann S.;
RL Submitted (JAN-2000) to the EMBL/Genbank/DBJ databases.
RN [5]
RP SEQUENCE OF 700-1239 FROM N.A.
RC TISSUE=Brain;
RX Saito T., Seki N., Hori T.;
RT "Isolation, expression profile and chromosome assignment of a novel
RT serine/threonine kinase gene.";
RL Submitted (Apr-1998) to the EMBL/Genbank/DBJ databases.

```

```

CC -1- FUNCTION: Serine/threonine kinase that may play a role in the
CC response to environmental stress and cytokines such as TNF-alpha.
CC Appears to act upstream of the c-Jun N-terminal pathway.
CC -1- CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphorylated protein.
CC -1- COFACTOR: Magnesium.
CC -1- SUBUNIT: Interacts with the SH3 domain of the adapter proteins
CC NCK (by similarity).
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=5;
CC Name=1; Synonym= tumor-associated;
CC IsoId=O95819-1; Sequence=Displayed;
CC Name=2;
CC IsoId=O95819-2; Sequence=VSP_007054, VSP_007057;
CC Name=3;
CC IsoId=O95819-3; Sequence=VSP_007056, VSP_007057;
CC Name=4; Synonym=HGK-S;
CC IsoId=O95819-4; Sequence=VSP_007054, VSP_007055, VSP_007057,
CC VSP_007058;
CC Name=5; Synonym=HGK-L;
CC IsoId=O95819-5; Sequence=VSP_007054, VSP_007055, VSP_007056,
CC VSP_007057, VSP_007058;
CC -1- TISSUE SPECIFICITY: Appears to be ubiquitous, expressed in all
CC tissue types examined. Isoform 5 appears to be more abundant in
CC the brain. Isoform 4 is predominant in the liver, skeletal muscle
CC and placenta.
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC STE20 SUBFAMILY.
CC -1- SIMILARITY: Contains 1 CNH domain.
CC -1- CAUTION: Ref.4 sequence differs from that shown due to a
CC frameshift in position 1151.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.jsb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; AF096300; AAD16137.1; -
CC EMBL; AY212247; AAC32626.1; -
CC EMBL; AB014587; BAA31662.1; -
CC EMBL; AL137755; CAB70907.1; ALT_FRAME.
CC EMBL; AB013385; BAA33714.1; -
CC Genew; HGNC:6866; MAPK4.
CC MIM; 604666; -
CC GO; GO:0005524; F:ATP binding; IDA.
CC GO; GO:0004674; F:protein serine/threonine kinase; IDA.
CC GO; GO:0004648; F:protein amino acid phosphorylation; IDA.
CC GO; GO:0007243; P:protein kinase cascade; IDA.
CC GO; GO:0007165; P:signal transduction; TAS.
CC InterPro; IPR001180; Cytro.
CC InterPro; IPR000719; Prot_kinase.
CC InterPro; IPR002290; Ser_thr_kinase.
CC Pfam; PF00780; CNH; 1.
CC Pfam; PF00069; pkinase; 1.
CC Pfam; PF00001; Prot_kinase; 1.
CC SMART; SM00036; CNH; 1.
CC SMART; SM00220; S_TKC; 1.
CC SMART; SM00219; TyKc; 1.
CC PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
CC PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
CC PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
CC ATP-binding; Serine/threonine-protein kinase; Transferase;
CC KW Alternative splicing.
CC DOMAIN 25 289
CC FT DOMAIN 844 1143
CC FT NP_BIND 31 39
CC FT BINDING 54 54
CC FT ACT_SITE 153 153
CC FT VARSPLIC 495 525
CC FT
CC /FTid=VSP_007054.

```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 4, 2003, 15:11:28 ; Search time 36 Seconds
(without alignments)
8694.939 Million cell updates/sec

Title: US-09-674-237A-3
Perfect score: 6269
Sequence: 1 MAQPTFGSGSIDVAITVE.....VGLPFSNVYKLTMDPDSQ 1213

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues
Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database :
- 1: SP archaea:*
 - 2: SP bacteria:*
 - 3: SP fungi:*
 - 4: SP human:*
 - 5: SP_invertebrate:*
 - 6: SP_mammal:*
 - 7: SP_mhc:*
 - 8: SP_organelle:*
 - 9: SP_phage:*
 - 10: SP_plant:*
 - 11: SP_rodent:*
 - 12: SP_virus:*
 - 13: SP Vertebrate:*
 - 14: SP_unclassified:*
 - 15: SP_virus:*
 - 16: SP_bacteriap:*
 - 17: SP_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the total being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4264.5	68.0	1220	13 Q8UFT5	Q8UFT5 brachydanio
2	4264.5	68.0	1721	13 Q8UFT4	Q8UFT4 brachydanio
3	1738.5	27.7	1014	5 Q8INU2	Q8INU2 drosophila
4	1735	27.7	1011	5 Q61639	Q61639 drosophila
5	1702	27.1	1097	5 Q9VIF7	Q9VIF7 drosophila
6	1698.5	27.1	1094	5 Q61618	Q61618 drosophila
7	1684	26.9	345	4 Q81WH9	Q81WH9 homo sapien
8	1530	24.4	611	4 Q81W15	Q81W15 homo sapien
9	1333	21.3	1097	5 Q9U2T9	Q9U2T9 caenorhabdi
10	1314	21.0	755	11 Q8C4B5	Q8C4B5 mus musculu
11	1172	18.7	440	11 Q8C9C3	Q8C9C3 mus musculu
12	822.5	13.1	164	11 Q8C1A3	Q8C1A3 mus musculu
13	812.5	13.0	158	11 Q8CGU5	Q8CGU5 mus musculu
14	794.5	12.7	662	5 Q8T068	Q8T068 drosophila
15	704	11.2	276	11 Q8CD59	Q8CD59 mus musculu
16	536.5	8.6	864	4 Q9UBC2	Q9UBC2 homo sapien

17	531	8.5	1253	5 Q9W111	Q9W111 drosophila
18	528	8.4	1253	5 Q8W061	Q8W061 drosophila
19	527	8.4	907	11 Q60902	Q60902 mus musculu
20	526.5	8.4	1232	5 Q8WMD2	Q8WMD2 drosophila
21	510.5	8.1	819	11 Q8CB70	Q8CB70 mus musculu
22	506.5	8.1	751	5 Q23418	Q23418 caenorhabdi
23	506.5	8.1	796	5 Q9B1F4	Q9B1F4 caenorhabdi
24	504.5	8.0	1106	5 Q8WMD3	Q8WMD3 drosophila
25	496.5	7.9	599	11 Q8CB60	Q8CB60 mus musculu
26	493.5	7.9	599	11 Q91WH8	Q91WH8 mus musculu
27	481	7.7	681	5 Q81FX4	Q81FX4 caenorhabdi
28	478	7.6	108	11 Q8CJ62	Q8CJ62 mus musculu
29	434.5	6.9	1116	5 Q9HGL2	Q9HGL2 schizosacch
30	405	6.5	944	5 Q81337	Q81337 caenorhabdi
31	405	6.5	1148	5 Q95Z77	Q95Z77 caenorhabdi
32	405	6.5	1439	5 Q8WQ7	Q8WQ7 caenorhabdi
33	405	6.5	1480	5 Q18138	Q18138 caenorhabdi
34	402.5	6.4	1003	12 Q91LX9	Q91LX9 kaposi's sa
35	397	6.3	1089	12 Q40947	Q40947 kaposi's sa
36	387.5	6.2	826	5 Q9VPS3	Q9VPS3 drosophila
37	386	6.2	1036	12 Q9DUM3	Q9DUM3 kaposi's sa
38	386	6.2	2376	5 Q9V5J0	Q9V5J0 drosophila
39	386	6.2	2376	5 Q966V1	Q966V1 drosophila
40	383.5	6.1	1909	5 Q25893	Q25893 plasmodium
41	382	6.1	1162	12 Q98148	Q98148 kaposi's sa
42	377.5	6.0	1129	12 Q9QR71	Q9QR71 kaposi's sa
43	375	6.0	1151	4 Q9BX65	Q9BX65 homo sapien
44	373	5.9	1596	5 Q81U44	Q81U44 plasmodium
45	367	5.9	976	12 Q9DUN0	Q9DUN0 kaposi's sa

ALIGNMENTS

RESULT 1

Q8UFT5 PRELIMINARY: PRT: 1220 AA.

AC Q8UFT5; 01-OCT-2002 (TREMBLrel. 22, Created)

DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)

DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)

DE SI:42173A8.1.2 (Novel protein similar to interseclin (SH3 domain protein, ITSN1)).

GN SI:42173A8.1.

OS Brachydanio rerio (Zebrafish) (Danio rerio).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;

OC Cyprinidae; Danio.

OX NCBI_TaxID=7955;

RN [(1)]

RP SEQUENCE FROM N.A.

RA Pandian R.;

RL Submitted (Aug-2002) to the EMBL/GenBank/DBJ databases.

CC -1- SIMILARITY: CONTAINS 5 SH3 DOMAINS.

DR EMBL: AL606751; CAD43427.1; -

DR InterPro: IPR002048; EF-hand.

DR InterPro: IPR000261; ERS15_homology.

DR InterPro: IPR00108; Neu_cyt_fact_2.

DR InterPro: IPR001452; SH3.

DR Pfam: PFO0036; efhand; 2.

DR	PRINTS: PR00499; P67PHOX.
DR	PRINTS: PR00452; SH3DOMAIN.
DR	PRODom: PD000066; SH3; 5.
DR	SMART; SM00054; EFh; 2.
DR	SMART; SM00027; SH3; 2.
DR	SMART; SM00326; SH3; 5.
DR	PROSITE: PS00018; EF_HAND; 2.
DR	PROSITE: PS00031; EF_2.
DR	PROSITE: PS00002; SH3; 5.
KW	SH3 domain.
SQ	SEQUENCE 1220 AA; 136351 MW; 1AB9134E2A4BA515 CRC64;
Query Match	68.0%; Score 4264.5; DB 13; Length 1220;

Best Local Similarity 68.3%; Pred. No. 1.5e-206;
Matches 869; Conservative 108; Mismatches 184; Indels 111; Gaps 21;

```

QY 1 MAQPTPPGSGSLDVAITVEERAKHDQFLSLKPL-AGHTGDDQARNRPFQSGLPQVLA 59
DB 1 MAQPTPPGSGSDTVIVISDERAKHDQFHSLSLTPAGITGDDQARNRPFQSGLPAPILA 60
QY 60 QMALADNNNDGRMOQVEFSIAMKLIKLIKLOGYOLPSTLPVWKQOPVALSAPAFGIG 119
DB 61 QMALADNNNDGRMOQVEFSIAMKLIKLIKLOGYOLPSTLPVWKQOPVALSAPAFGIG 108
QY 120 IASMPDLTAVALPVGMSIPVVGMSPPVSVPPAAVPPPLANGAPVIOPLPAFAHPAATW 179
DB 109 --STP-----PTPG-VPGMGVSPPLGASVTP-PVPSLANGAPAMIDPLGFSHGAVL 159
QY 160 PKSSFSFGSGPSQNTKLQKQAFDVAS--APPAEAHVQSSRLKTYROLFNSHDKTM 236
DB 160 NKTSPFNRS-----SILKQKQSVAPAPAPAPADMAVQSSRLKTYROLFNSHDKTM 213
QY 237 SGHLGPOARTILMQSSLPOAQLASTIMNLSIDIDQGLTAEEFILAMHLIDVAMSGOPLP 296
DB 214 SGYLTGPOARTILMQSSLPOAQLASTIMNLSIDIDQGLTAEEFILAMHLIDVAMSGOPLP 273
QY 297 PVLPEPYIPPSFRRVRSQSGMSVSSSVYDORLPPEPSSEBDOQPEKKLPVTFEDKKEN 356
DB 274 PLPLPDLIPPTFRMRSGSVSVTSMHSTDLASQEEP-EBEKEAKKKLPVTFEDKKEN 332
QY 357 PERGSVELKRRQALLBOQRKEQERLAQERAEQERKEREQERQERQERQERQERQER 416
DB 333 PERGMLEKRRQALLBOQRKEQERLAQERAEQERKEREQERQERQERQERQERQER 392
QY 417 ELEROREBERREKREIRREAKKELEROROLEMERRROROLENQRKEQEGTVLAKRKT 476
DB 393 ELEROREBERREKREIRREAKKELEROROLEMERRROROLENQRKEQEGTVLAKRKT 452
QY 477 LEFELALNDKKQLEGKLDIRCLATOROIEESTNKSRELIAETHLQOOLQESQOM 536
DB 453 LEFELALNDKKQLEGKLDIRCLATOROIEESTNKSRELIAETHLQOOLQESQOM 512
QY 537 LGRLLPEKQIISDLQKOVQNSLHSDSLTLTKALAEKELAQOLREQLDEVERTSKL 596
DB 513 LSRLIPDKQCLNDQKOVQNSLHSDSLTLTKALAEKELAQOLREQLDEVERTSKL 572
QY 597 QEIDVFNNQKELREIHSKQOLQKORSLEAARLKKOEKRSLELEKQEDAQORVQERD 656
DB 573 LEIDAFNQKELREIHSKQOLQKORSLEAARLKKOEKRSLELEKQEDAQORVQERD 620
QY 657 KQMLEHVQOQEQP-----RPRKPHEDRLKQEDSVRKK--EAEERAKPEMQDKQS 704
DB 621 --LSGTDGVSFAWRDDGLGKAPRPPVQAMWSR--VREKKIVDAENHGFQCCCFM 674
QY 705 RLPHHQBPAKLAQAPMSTTEKQPLTSAQSVKVVYRRLALYPRESKSHDITIQPDI 764
DB 675 RVNVRNQHNSCVC-----VLDKIPVSGFNQKVVVVYRRLALYPRESKSHDITIQPDI 728
QY 765 VAVDSSQCEPGLGELGKTGWFPANAEKIPENEVPTPAKPYTDLTSLA----- 816
DB 729 VAVDSSQCEPGLGELGKTGWFPANAEKIPENEVPTPAKPYTDLTSLA----- 788
QY 817 -----APKALRETPAPL-PVTSSEPTP-----NNMADFSSSTWPSNS-EXPE 859
DB 769 ASSTTATATILPVSTEPASIAPPSSAPPTGPASSSSSSASMSWADPSTWPSNSAVEKD 848
QY 860 TNNMTW---AAQPSITVPSAQOLQKRSAPATATGSSPSVVLQGEVVEGLQALALP 916
DB 849 SGGWAMPQTPQPSLSPSGQIRORSAPATATGSSPSVVLQGEVVEGLQALALP 908
QY 917 WRAKDDNHLNFKSDVITVLEQDDMMWFGEGVGOGMFKSVYKLSGVPKSTSIDGP 976
DB 908 WRAKDDNHLNFKSDVITVLEQDDMMWFGEGVGOGMFKSVYKLSGVPKSTSIDGP 968
QY 977 TESPAFLKRVASPAKPA-----IPGEEFIAMTYESSEQGLTFOQGDVIV 1023
DB 977 TESPAFLKRVASPAKPA-----IPGEEFIAMTYESSEQGLTFOQGDVIV 1023

```

```

DB 969 SDSPPSVKRPSPINLKTTELCEGQNSNSNMYPSKEYVAMTYESNEQGLTFOQGDVIT 1028
QY 1024 VTKKDGMMWCTVQDCKSGVPSNRYRLKDSQSGCTACKTSGIKKPRIAQVIASTATGP 1063
DB 1029 VTKKDGMMWCTVQDCKSGVPSNRYRLKDSQSGCTACKTSGIKKPRIAQVIASTATGA 1068
QY 1084 EQLTLARQOLILIRKKRPGWMEGELQARQKQIGMFPANVYVLLSPGTSKITPTE--L 1141
DB 1089 EQLTLARQOLILIRKKRPGWMEGELQARQKQIGMFPANVYVLLSPGTSKITPTE--L 1146
QY 1142 PKTAVPACVIGMYDTYATNDELAFPSQOIIVINKEDPDWKKGEVSGQVGLPFSNY 1201
DB 1149 PKLTPPAAVCCVIGMYDTYATNDELAFPSQOIIVINKEDPDWKKGEVSGQVGLPFSNY 1208
QY 1202 VKLTMDPSQO 1213
DB 1209 VKLTMDPSQO 1220

RESULT 2
Q8JFT4 PRELIMINARY; PRT; 1721 AA.
AC Q8JFT4
DT 01-OCT-2002 (TREMblrel. 22, Created)
DT 01-OCT-2002 (TREMblrel. 22, Last sequence update)
DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
DE SI.dz173a8.1.1 (Novel protein similar to Intersectin (SH3 domain protein, ITSN1)).
GN SI.dz173a8.1.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
CX NCBI_TaxId=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Pandian R.;
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC -1- SIMILARITY: CONTAINS 5 SH3 DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 C2 DOMAIN.
DR EMBL; AL606751; CAD43428.1; -.
DR InterPro; IPR000008; C2.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR000261; EPS15 homology.
DR InterPro; IPR00108; Neu_cyt_fac2.
DR InterPro; IPR001849; PH.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00168; C2; 1.
DR Pfam; PF00169; PH; 1.
DR Pfam; PF00621; RhGEF; 1.
DR PRINTS; PR00499; P67PHOX.
DR PRINTS; PR00452; SH3DOMAIN.
DR ProDom; PD000066; SH3; 5.
DR SMART; SM00239; C2; 1.
DR SMART; SM00054; EFh; 2.
DR SMART; SM00027; EH; 2.
DR SMART; SM00233; PH; 1.
DR SMART; SM00325; RhGEF; 1.
DR SMART; SM00326; SH3; 5.
DR PROSITE; PS50004; C2 DOMAIN_2; 1.
DR PROSITE; PS50010; DH_2; 1.
DR PROSITE; PS50018; EF_HAND; 2.
DR PROSITE; PS50031; EH; 2.
DR PROSITE; PS50003; PH DOMAIN; 1.
DR PROSITE; PS50002; SH3; 5.
KW SH3 domain.
SQ SEQUENCE 1721 AA; 194115 MW; 2857E789584D9C07 CR664;

Query Match 68.0%; Score 4264.5; DB 13; Length 1721;
Best Local Similarity 68.3%; Pred. No. 2.3e-206;

```

Matches 869; Conservative 108; Mismatches 184; Indels 111; Gaps 21;

```

Qy 1 MAQPTFGSLDVAITVEERAKHDQQLSLKPI-AGFITGDOANFPFGSLPOPVLA 59
Db 1 MAQPTFGSGSDTWISYDERAKHDQSHSLTPPAGITTDQANRFLQSLGLPPIIA 60
Qy 60 QIUALADNNDRMDQVEFSIAMKLIKLOQYOLPSTLPVWKQOPVAISSAPAGITG 119
Db 61 QIUALADNNDRMDQVEFSIAMKLIKLOQYOLPSTLPVWKQOPPI----- 108
Qy 120 IASMPPLTAVAVPMKSIPIVNGKSPVSVPPAANVPLANGAPPIQOLPAPAPAAAT 179
Db 109 --SIP-----PTPFG-VPMGVSPPLAGSVTP-PPVSLANQAPAIQIPLSGSHGAVL 159
Qy 180 PKSSPFRSGPSQSLNTKLOKXASFPVAS---APPAEAVAPQSSRLKTROLFNSHDKTM 236
Db 160 NKTSPFNRS-----SILQKQGVAPAPAPAPADNAVSSSLKTROLFNSHDKTM 213
Qy 237 SGHLTGPOARTILMOSLTPQAOLASIMWLSLIDIDQDKLTAEFFILAMHLIDVAMSGQPLP 296
Db 214 SGYLTPQARTILMOSLTPQAOLATIMWLSLIDIDQDKLTAEFFILAMHLIDVAMSGPLP 273
Qy 297 PVLPEYIPPSFRVAVSGSGMYISSVYDQRLPEPSSSEDDQOPKPLPTPEDKREN 356
Db 274 PLPLPDLIPPTFRMRKSGSVSTVMSHTDLSQEEP-EEEKKAKKPLPTPEDKREN 332
Qy 357 FERGVELKRRQALLQOKRQOERLAQLEBAOEKREKREBOEKROLELEKOKER 416
Db 333 FERGVELKRRQALLQOKRQOERLAQLEKOEKREKREBOEKROLELEKOKER 392
Qy 417 ELEROREERREKREERREAKRELEOROLEWERNRQELLNQNKQOEGTVLAKARKT 476
Db 393 ELEROREERREKREERREAKRELEOROLEWERTROQLLNQNKQOENIVLAKARKT 452
Qy 477 LEPELEALNDKXQLEGLQIDTRCRLATROEITESTNKREIRIATITLQOOLQESQOM 536
Db 453 LEPELEALNDKXQLEGLQIDTRCRLATROEITESTNKREIRIATITLQOOLQESQOM 512
Qy 537 LGRLLPEKQILSDQKQVQONSILHRPSLLTLKRALEAKELAQOALEQOLDEVERERSTL 596
Db 513 LSRLLPDQKQNDQKQVQONSILHRPSLNSLOALIMKSTYQOOLEQOLDAVEKRETRSL 572
Qy 597 QEIDVFNQKLEKEIRHSKQOLQOKRSLBAARLKQOEKREKSLTEKOKEDAQORQERD 656
Db 573 LEIDAFNQKLEKEIRHSKQOLQOKRSELE-ADITQSHRKSLSDSDSR----- 620
Qy 657 KQWLEHVQOEEP-----RPRKHEEDRLKREDSVRK--BAERAKEMQDKOS 704
Db 621 ---LSGTDGVPAMWDGIGKAPRPPVSGAMWSR---VREKKIVDAERHGFQCCCFM 674
Qy 705 RLPHHQEPAKLATAQPMSTTEKGLITIAQESVKKVYVYRALYPPFSRSHDETTOPGDI 764
Db 675 RVNVRQKHSQVC-----VLNDIPIVSGNQEKKVYVYRAMVPFARSHDETTIHPGDI 728
Qy 765 VNVDESQTEPGLGELKGTGMPFANYAEKIPENEVETPAKPVYDLTISAP----- 816
Db 729 VNVDESQTEPGLGELKGTGMPFANYAEKIPESVPLSLRASAASSAPRLGSHMS 788
Qy 817 -----APKALRETAPL-PVTSSEPTP-----NNWADFSSTWPSSSN-EKPE 859
Db 789 AASSTTATPILPVSTEPASIAPPSSAPPTGPASSSSASSANWADPSTTWSNAVEKQD 848
Qy 860 TQNWQD---AAQPSLTVPSAGOLRQSAFTPATATGSSPVLGGGEKVEGLOQALYLP 916
Db 849 SGQWMDMPQTPQPSLTVPSGQIRQSAFTPATITLSSSPSPVLGGGEKVEGLOQALYLP 908
Qy 917 MPKAKDNHILFNKSDVITVLEQODMMWFGVQOKMFPKSYKLTJSGPVKSTSIDTGP 976
Db 909 MPKAKDNHILFNKSDVITVLEQODMMWFGVQOKMFPKSYKLTJSGPVKSTSIDTGP 968
Qy 977 TSPSPALKEVASPAKPA-----IPGEFIAMVYTESQGLTFQOGDVIV 1023
Db 969 SDSPSPVKKRPSILNKPTLGEQGNNSNMYSPKEYVAMVYTESQGLTFQOGDVIV 1028

```

Qy 1024 VTKKGDMMWTGTVGDKSGVFPSSNVYRLKDSBSSGTAGTGLSGKKPEIAQVIAAATGP 1083
 Db 1029 VTKKGDMMWTGTVGDKSGVFPSSNVYRLKDSBSSGTAGTGLSGKKPEIAQVIAAATGP 1086
 Qy 1084 EQLTLAPQQLIIRKKNPGQWEGELQARKKRQIGWFPFANYKLLSPGTSKITPTE--L 1141
 Db 1089 EQLTLAPQQLIIRKKNPGQWEGELQARKKRQIGWFPFANYKLLSPGTSKITPTE--L 1148
 Qy 1142 PRTAOPAVCCVIGWYDTAQNDELAFSGQIIVLVKKEPDMKGEVSGQVGLPSPNY 1201
 Db 1149 PRLPFPNVCQVIGWYDTAQNDELAFSGQIIVLVKKEPDMKGEVSGQVGLPSPNY 1208
 Qy 1202 VKLTTDMPDPSQ 1213
 Db 1209 VKLTTDTPDSQ 1220

RESULT 3
 08INU2
 ID 08INU2; PRELIMINARY; PRT; 1014 AA.
 AC 08INU2;
 DT 01-MAR-2003 (Tremblrel. 23, Created)
 DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
 DE CG1099-PB.
 GN DAPI60.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.H., Blazej R.G., Champe O., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
 RA Abril J.F., Agbayani A., An H.J., Andrews-Pfankuch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktoglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Bertman B.P., Bhandari D., Bolshakov S.,
 RA Borzova D., Botchan M.R., Bouck J., Brokstein P., Brotier P.,
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Folsler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasco P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mallet B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Mervilov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D., Scheefer F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski P., Smith T.,
 RA Splitter R., Tector A.C., Stapleton M., Strong R., Sun E.,
 RA Svitek A.R., Turner C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.Y., Wasserman D.A., Weinstock G.M., Weisenbach J.,
 RA Williams S.M., Woodgett M., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
 RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195 (2000).

[2]
 RP SEQUENCE FROM N.A.
 RA Celniker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
 RA Batzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Bisam D.A.,
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
 RA Dodson K., Dorett V., Doup L.E., Doyle C., Dresnek D., Fartan D.,
 RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
 RA Ibegwam C., Jallil M., Kruse D., Li P., Maitel B., Moshrefi A.,
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nuno J.,
 RA Paciel J., Parasas V., Park S., Patel S., Pfeiffer B.,
 RA Phouanavong S., Pitman G.S., Puri V., Richards S., Scheeler F.,
 RA Stepieton M., Strong R., Svirskas R., Tector C., Tyler D.,
 RA Williams S.M., Zaveri U.S., Smith H.O., Venter J.C., Rubin G.M.,
 RT "Sequencing of Drosophila melanogaster genome."
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Miera S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
 RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
 RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
 RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
 RA Krommiller B., Marshall B., Millburn G., Richter J., Russo S.,
 RA Searle S.M.J., Smith E., Shu S., Smutnick F., Whitfield B.,
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.,
 RT "Annotation of Drosophila melanogaster genome."
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter J.C.,
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RA FlyBase;
 RL Submitted (SEP-2002) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AE003669; AAN11099.1; -;
 SQ SEQUENCE 1014 AA; 111890 MW; 633790A316885FB4 CRC64;

Query Match 27.7%; Score 1738.5; DB 5; Length 1014;
 Best Local Similarity 33.5%; Pred. No. 1.7e-79;
 Matches 414; Conservative 204; Mismatches 332; Indels 285; Gaps 31;

QY 11 SLDVWATIVERAKIDQOPLSLKPIAGITGDQANFPFQSGLPVLAQIWMALDMND 70
 DB 4 AVDMAAVPRERLKYQEOFRALQPAQGVTAQAQGFPLQSGLPPLIIGQIWMALDITSD 63
 QY 71 GRMDOVERSIAMKIKLKLQGYOLPSTLPVWKQOPVAISSAPAFGIGIASMP---LT 127
 DB 64 GDMINERISACKLNLKRGMDVFKVLPSSL-----LSSL-----TEDVFSMTFRGSTS 113
 QY 128 AVAPVPMKSIPIVNGSPPLVSSPPPAVPLPLANGAPVYIOPLPAPAPHAATWPKSSFSR 187
 DB 114 SLSPLD---PLKGI-VPAAVAVPVVAPVVA--VATVISP-PGVSVSPSGFPPTSN--- 162
 QY 188 SGPGSLNTKLOKQASFDVAAAPPAEWAIVPOSSSLKTRQLFNSHDKTMSGHLTPQART 247
 DB 163 --PFRHRTSISERAPSISISVN---OGEWAVQAQKRTQYVFNADRTSRGTLTSQARG 217
 QY 248 ILMOSLQOAOALASITWNLSDIDQDKLTAAEFITLMHLDIVMSPGQIPVLYPREYIPPS 307
 DB 218 VLVQSLKPOVTLAQITLSDIDGRNLNDEFITLMFLCEKRAAEKGIPTVLPQEWVFPN 277
 QY 308 FRRVSGSGMSVYSSSVDOPLREPPSSD-----EQQPEKLP--VTFEDKK 353
 DB 278 LRKISRG---SVSGVVSRRGSPASHASVSSQSGVVDADPTGLPQGTSEFDKX 333
 QY 334 RENFRGSGVELEKRRQALIEQKQKQERLAQLEFRAEQEKERERQOAKQLELEKQLE 413
 DB 334 KENYVGAQELDRRRKIMEDQOKERERERERERERERERERERERERERERERERER 393
 QY 414 KORELEROREERREERREERREERREERREERREERREERREERREERREERREER 473

DB 394 ROREIEMEKEQRRKLEAKREARKELEKORQOEWEQARIENMAQREERQEVHLKOKAH 453
 QY 474 RKLFELELNLNKQQLBEKQDIDICRLATQGOEISTNKRELRILAEITHLQOQOES 533
 DB 454 NTQNLWELSTLNKIKELSORICDTRAGVNVVTVIDGMETQDTSSEWSQLARKIQ 513
 QY 534 QOMGLRIPK-----OILSPQLQVOONSILHRDSTLTLKALBAKELARQOLR 582
 DB 514 NAKLQUTQORAKREAKSKASGAALGEMNAQOE-----LMAFAHKOILINQIK 563
 QY 583 EOLDEVERETRSKLOEIVENNOLKEIRE-----HSKQOLQORSLEAARLKQEKERK 637
 DB 564 DKVENISKETIESKEDINTNDVQMSKELASLITCEPLLYKXYDQRTSVLELKNRK 623
 QY 638 SLEEKKEQAKRVRQORDQWLEHVQOEOPRPRKHEBDRLKREDSVRKKEAKERAKP 697
 DB 624 NETSVSSAMPT-----GSSSAW-----EEGTVTYDYY-----AVASNDLSALAAP 664
 QY 698 EMODKQRLPHPHQEPAKLTAQAPWSTTEKGPLTISAQESVYVYALYPFESRSHDEI 757
 DB 665 AVD-----LGGPAP-----EG-----FVKQAVVEFNARNAEI 693
 QY 758 TIOPGDIVMWDESGTQEPGMLOGLKQKQWFPANVAEKIPENEVPTPAKPVTDLTAPA 817
 DB 694 TFPVGDITLVLPEQNAEPGMVLAGEINGHTGMFPESYVEKLEVGCV-----A 739
 QY 818 PKALRETPAPLPVTSSPESTTPPNWADPSSSTWSSNEKPEPDNDMTMAQPSLTVPSA 877
 DB 740 PVAAVE---APVDAQVATVADTVN-----DNINT-----SSIPAA 771
 QY 878 GOLRQSAFPAATATGSSPSPVLQCGEKEVGLQALYPRAKKNHLPNFKSDVITVLE 937
 DB 772 S-----ADLTAG-----DV----- 781
 QY 938 QQDMWFEVQOGKWPFPKSYVKLISGPVRKTSIDTGPTEPSASLRVASPAKPAIPG 997
 DB 782 ----- 781
 QY 998 EEFIAMTYESSEGDITFQOQGVIVYTKKDDGMWTTQVDKSGVPSNRYRLKDSBSG 1057
 DB 782 EYVIAAPYPSAEAGDISFSAGENVWVYKKEGEMWTGTSRGTGMPFSNVYQKADVGTAS 841
 QY 1058 TAG-----KTSLGKKPEIAQVIASVYATGPQTLTAPQGLILIRKKNPGWMEGELQ 1110
 DB 842 TAAEPESPESLDQGRARAKSEIAQVIAYEATSTQLSLTGQILIMIKKTDGSMWBELO 901
 QY 1111 ARGKKRQIGWFPANVYKLLSPG--TSKITPTLEPKTAVQPAVC--QVIGMTDYTAQNDEL 1167
 DB 902 AKGRRQIGWFPATYVAVLQGRNSGRNTPVSGSRIEMTEQILDKVIATLYYKAQNDEL 961
 QY 1168 AFSKQIINVLNKEDPDMWKGVEVGQVGLPPSNV 1202
 DB 962 SFDKDDIISVLGRDEPEWMRGELNGLGLFESNV 996

RESULT 4
 ID 061639 PRELIMINARY; PRT; 1011 AA.
 AC 061639;
 DT 01-AUG-1998 (Tremblrel. 07, Created)
 DT 01-AUG-1998 (Tremblrel. 07, Last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
 DE Dynamically associated protein isoform DAPI60-2.
 GN DAPI60 OR CG1099.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydridae; Drosophilidae; Drosophila.
 OK NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=OREGON-R; TISSUE=Head;
 RA Roos J., Kelly R.B.,

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpén G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kratke C., Kravitz S., Kulp D., Lai Z.,
RA Laebo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mettel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milhina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Pacleb J.M.,
RA Palazzo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheller F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier B., Spreading A.C., Stapleton M., Strong R., Sun E.,
RA Svrtsek R., Tecor C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of *Drosophila melanogaster*.";
RL Science 287:2185-2195 (2000).
CC -1. SIMILARITY: CONTRAINS 4 SH3 DOMAINS.
DR EMBL; AE003669; AAF53962.1; --
DR HSSP; P29354; IGRI.
DR FlyBase; FBgn0023388; Dap160.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR000261; Eps15_homology.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00036; ehfand; 2.
DR Pfam; PF00018; SH3; 4.
DR PRINTS; PR00452; SH3DOMAIN.
DR ProDom; PD000066; SH3; 4.
DR SMART; SM00054; Eph; 2.
DR SMART; SM00027; EH; 2.
DR SMART; SM00326; SH3; 4.
DR PROSITE; PS00018; EF_HAND; 2.
DR PROSITE; PS00031; EH; 2.
DR PROSITE; PS50002; SH3; 4.
DR KEGG; Calcium-binding; SH3 domain.
SQ SEQUENCE 1097 AA; 120766 MW; 2573D7F8DCBACB CRC64;

Query Match 27.1%; Score 1702; DB 5; Length 1097;
Best Local Similarity 31.7%; Pred. No. 1.3e-77;
Matches 418; Conservative 203; Mismatches 329; Indels 368; Gaps 33;

11 SLIDVAIVTEERAKDQOFLSLKPTAGFTGQANFPFQSGLPVLAQIMALDMMND 70
4 AVDAVAIVPRERLKYOEQFRALQPOAGFYGAQAQFLLQSLPLILGQIWAALDTSDD 63
71 GRMDVRSIANKLTKLKLQGYLPSTLPVVKQOPVAISSAPFGIGIASMP---LT 127
64 GKMNINSEIACKLNLKLRGMDVFKVLPSSL-----LSSL-----TGDVRSMTRGSTS 113
128 AVAPVPMGSIPIVVGMSPLVSSVPAAVPPPLNGAPVYIQLPAFAHPAATWPKSSFSR 187
114 SLSPLD-----PLKGI-VPAVAVPVVAVPVA--VATYISP-IGVSVSGPFPPTSN--- 162
188 SGFGSLTKLQKQAFDVAAPPAEAVAVPOSSSLKTRQLFNSHDKTMSGHLTPQART 247
163 --PSPRHTSISERRAPSIRESV---QGEWAVQAQKRYTQVFANADRTTRSGYLTSSQARG 217
248 ILMGSSLPQAOLASIMNLSDDQGLTAEEFILMHLIDVMSGQPLPVVLPREYIPPS 307
218 VLVGSKLTPQVTLAQIWLTLSDIDGRNLNDEFILMLFCEKMAQAEKLPVTLPOEHWPPN 277
308 FRFVSGSGMSVSISSSVQRLPEPSSSD-----EQQPEKLP--VFEDBK 353
278 LRKISRG-----SVSGVSRSGSQPASHASVSSOSGAVVDAPRTAGLPQSTFEDXR 333
354 RENFSGVLEKRRQALLIEQQRKEQLAQLERAEQERKERQEQEAKRLBLEKOLE 413
334 KENYVGAQELDRRRKIMEDQQRKERERERERERERERERERERERERERERERER 393
414 KQRELERQERERERERERERERERERERERERERERERERERERERERERERERER 473

394 RQRELEWEKEQRERERERERERERERERERERERERERERERERERERERERER 453
474 RKLTEFLEALNDKQKQLEGLQDIPRLATQROEISTNSKRLRLAETHLOOQOES 533
454 NTQNLVELSTLNKEIKELSORICDTRAGVTVKTVIDGMTQRTQRTSSEMSQLARIKEQ 513
534 QOMGLRLEPEK-----QILSDQLKQVOQNSLRSDSLTLKRLALEKELELARQOLR 582
514 NAKLQLTQERAKKAKSAGSALGEMNAQOE-----LNAFAHKLIIINQIK 563
583 EQLDEVERETRSKLQEIIVFNNOKELE-----HNSKQOLQKRSLEARLKQEQERK 637
564 DKVENISKEIESKEDINTMDVQMSLKAELSLATICEPDIKEYDVQRTSVLELKNRK 623
638 SLELEKQEDAKQRRVQERDKQMLEHVQOEOPRRKHEEDRLKREYSKKKEBEQAKP 697
624 NETSVSAMPT-----GSSSNW-----EEGTITVDY-----AVASNDISALAAP 664
698 EMQDKQRLPFPHQEPAKLATQAPWSTTEKQPLTISAQESVYVYRALVPFESRSHDEI 757
665 AVD-----LGAPAP-----EG-----FKYQAVIEFNARNAEIEI 693
758 TIQPGDIWMYDESQSGPGLGELKGTGMPANVAEKIPENEVPTPAKPVTDLTSAQA 817
694 TFVPGDIIVPLEGNAPGWLAGEINGHTGMPFESYEKLEGEV-----A 739
818 PKALRETPAPLPVTSSEPTTPPNMADFSTTSSSNEKETNNMTMAAQLTYVPSA 877
740 PVAAVE---APVDAQVAVTDVTVN-----DNINT-----SSIPAA 771
878 GQLRQRAFPTATGSSPSPVLQGEKVEGLQAQALYPRAKKDNHNFKSDVITYLE 937
772 S-----ADLTAG-----DV----- 781
938 QQDMMWFGEVQGGKMFPSKYVKLISGPVKSTSIDTPTESPASLKRVASPAKPAIPG 997
782 ----- 781
998 EEFIAMTYTESSEGGDLTFQGDVIVYTKDQDMWTGVDKSGVPSNRYRLKD----- 1052
782 EYVIAAPYSEAEAGDLSFSGEMVWVYIKKEGEMWTGITSRTGMPFSNVYQKADVGTA 841
1053 -----SESG----- 1057
842 TAAAEVPSLDOETTLNGNAATYAPVBAQVYQIPVQEPSPROPISRGVGAEEAHED 901
1058 -----TAGKT-----GSLGKKPEIAQVYASVYATGPQOLT 1087
902 LDTESQINQSKTQOSEPAESYSRPMSTRSSMTPGRAKRSLEIAQVIAPEATSTQOLS 961
1088 LAPQQLILIRKKNPGWMEGELQARKKRRQIGMPANVYKLLSG--TSKTPLELXKA 1145
962 LTRQQLIMIRKKTGSGWMEGELQARKRRQIGMPANVYKLLSG--TSKTPLELXKA 1021
1146 VQPAVVC-QVIGMYVYTONDELAFSGQIIVNLNKEDPDMWKGVEGQVGLPSPNVY 1202
1022 MTEQILDLVIALYFYKQONDELSTFDODIISVIGRDEPPEMRGELNGLSGLPSPNVY 1079
RESULT 6
ID 061618 PRELIMINARY; PRT; 1094 AA.
AC 061618;
DT 01-AUG-1998 (TREMblrel. 07, Created)
DT 01-AUG-1998 (TREMblrel. 07, Last sequence update)
DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
DE Dynamn associated protein isoform DAP160-1.
GN DAP160 OR CGI1099.
OS *Drosophila melanogaster* (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Insecta; Pezomyzota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; *Drosophila*;

NCBI_Taxid=7227;
 (1)
 RN SEQUENCE FROM N.A.
 RC TISSUE=Head;
 RA Roos J., Kelly R.B.;
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: CONTAINS 4 SH3 DOMAINS.
 DR EMBL; AF053957; AAC39138.1; -
 DR HSSP; P29354; 1GR1.
 DR FlyBase; FBgn0023388; Dap160.
 DR InterPro; IPR002048; EF-hand.
 DR InterPro; IPR000261; EFS15_Homology.
 DR InterPro; IPR001452; SH3.
 DR Pfam; PF00036; efhand; 2.
 DR Pfam; PF00018; SH3; 4.
 DR PRINTS; PR00452; SH3DOMAIN.
 DR PRODOM; PD000066; SH3; 4.
 DR SMART; SM00054; EFh; 2.
 DR SMART; SM00027; EH; 2.
 DR SMART; SM00326; SH3; 4.
 DR PROSITE; PS00018; EF_HAND; 2.
 DR PROSITE; PS00031; EH; 2.
 DR PROSITE; PS00002; SH3; 4.
 KW SH3 domain.
 SQ SEQUENCE 1094 AA; 120495 MW; 141370EB1FB960F0 CRC64;

Query Match 27.1%; Score 1698.5; DB 5; Length 1094;
 Best Local Similarity 31.6%; Pred. No. 1.9e-77;
 Matches 417; Conservative 201; Mismatches 329; Indels 371; Gaps 33;

11 SLDWATVTEERAKHDQFLSKPIAGFITGDQARNFPGSGLPQVLAQIWMALADNMND 70
 4 AYDAMAVTPREERAKYEOGFRAIQPOKGFVGTGAQAKFFLQSQPLILQIWMALADTSD 63
 71 GEMDQVEFSIAMKLILKLTQGYQLPSTLPVPMKQQAISAPAFGIGIASMP--LT 127
 64 GKGNINIEFSIACKLIMIKLGMQVPRVLPSTL--LSSL--TGDPVSWTPRGSTS 113
 128 AVAPVMSGISPVVSGMPLVSSVPPAAVPPPLANGAPVYQPLPAFPAHPAATPKSSFR 187
 114 SLSPDLD---PLKGI-VPAVAPVPPVAPVPA--VATVISP-FGVSVPFGPTPTPTSN-- 162
 188 SGPGSGLNTKLOKAFSDVASAPPAEMAVNPQSSRLKYROLFNSHDKTSGHLTGPQART 247
 163 --PSPHNTISERAPIESVN--QGMNAVQAOKKKTQVRNANDRTSGVHTSGQANG 217
 248 IIMQSSLPQAOLASINWLSIDIDODGKLTAEFFILAMHLIDVAMSGQPLPPLPPEYTPS 307
 218 VLVQSYLPQVTLAQIWTLSIDIDODGRINCDEFILAMFLCEKMAWGEKIPVTLPQEWVPPN 277
 308 FFRVNSGSGMSVYSSSSVQRLPEPSSSD-----EQPEKLP--VTFEDK 353
 278 LKKIKSRP-----SVSGVVSPPSQSPASHVASVSSQSGVGVADADPTAQLPGQTSFEDK 333
 354 RENFERGSVELEKRRQALTEQOKKEOERLAOLEBAROEKREKREKROEAKOLELEKOLE 413
 334 KENYVQGAELDRKKIMEDQOKKEBEERKEEREAADKREKRLAEKQOEELRQO 393
 414 KORELEKREERKEERKEEREAARKELEBOROLEMRNRROELNQRNKEOGSTVVLKAR 473
 394 ROREIEMERKEORKELEKARKELEKOROEWEQAIEMNAQKEREGERVILKQXAH 453
 474 KRTLEFEALNDKQOLEGKLODTCRLATROEIESTNKSRELIATITLQOOLQOS 533
 454 NTQNLVELSTLNEKIKELSORICDTAGVNTVTDGKRTORDTSSMSQKAKIKQO 513
 534 QOMAGRLIEK-----QILSDQLKOVQONSILHRDSLTLTKALEAKELANQQR 582
 514 NAKTLQLOTERAKWEAKSKASGALAGENAQOEQ-----LNAAPAHQOLIINQK 563
 583 EQLDEVERETREKLOEIDVFNQOLKELR-----IHSKOOLOKORSLEAARLKQOEK 637
 564 DKVENISKEISKEDINTNDVQMSSELKALSLALITKCDLYVEYVGVQTSVLELKNK 623

638 SLELEKQEDARROVERDKQWLEHVQOEOPRPRKPHBEDRLKREDSYRKKEABERAKP 697
 624 NETSVSSAMDT-----GSSSAW-----BETGTTTDPY-----AASNDISALAP 664
 698 EMQODKOSRLFHPHCPAKLATQAPWSTTEKGPLTISAOSVYVYRYALYPRESHDEI 757
 665 AYD-----LGGPAP-----EG-----FVYQAVYENNAABEI 693
 758 TIQPGDIYMWDESQIGEPGMAGELKGTGMFPANYAEKIPENEVPTPAKPVTLTSDAP 817
 694 TEVPEDIIILVPLEQNAEPQWLAGINGHTGMFPESYVEKLEVEGV-----A 739
 818 PKLARETPAPLPVYSSSEPTTPNNWADPSSSTWPESSNEKPEPTDWMQMAQPSLTVPSA 877
 740 PVAAV-LEAPVDAQVADT-----YNDNINT-----SSTIPA 768
 878 GQLRORSAFTPATATGSSPSPVLCQGEKEVEGLQALYPMWAKDNHNFNKSVDITVLE 937
 769 S-----ADLTAAQ-----DY----- 778
 938 QODMMWFGEVQCGKMPKSYVKLISGVRKSTSIDTPTSPASLAKRVASPAKPAIFG 997
 779 ----- 778
 998 EEFIAMTYESESQGLTFQCGDVIATVKKQDMWTGTGVGKSGVFPSPNYVRKLD----- 1052
 779 EYIAYAPYESAEBGLSFSAGEMWVIKKEGEMWTGTISRTGMFPSPNYVQKADVTAS 838
 1053 -----SESG----- 1057
 839 TAAAPVESLDQETTLNNGAAYTAPVPAQEVYQPLVQERSBPISPGVGAABEAD 898
 1058 -----TAGKT-----GSLGKPEIAQVIASVYATGPEOLT 1087
 899 LDETVSQINTQSKTQSSPAESYSRPMGRTSMTPGMAKSEIQAIVLAPYATSTEDLS 958
 1088 LAPGOLILIRKNPGGMWEGELQANGKRRQIGFPAANYKLLSPG--TSKITPTLPKTA 1145
 959 LTRGOLIMIRKKTDSGMWEGELQANGRRRQIGFPAATVYKVLQGGRRNSGRNTPVSGSRIE 1018
 1146 VQPAVC-QVIGMYDVTQAQNDDELAFPSKQIINVLNKEDPDMMWKGESGOVGLFSPNYV 1202
 1019 MTEQLIDKVIATLPYKQANDDELSPDKDIIISVLRDEPMMWKGELNLSGLFSPNYV 1076

RESULT 7
 Q81WH9 PRELIMINARY; PRT; 345 AA.
 AC Q81WH9;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Similar to interectin 1 (SH3 domain protein) (Fragment).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 ON NCBI_Taxid=9606;
 RX (1)
 RC SEQUENCE FROM N.A.
 RP TISSUE=uterus;
 RA Struhsberg R.;
 RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC039036; AAH39036.1; -
 FT NON TER 345
 SQ SEQUENCE 345 AA; 37077 MW; 967527FDE570DE85 CRC64;

Query Match 26.9%; Score 1684; DB 4; Length 345;
 Best Local Similarity 95.3%; Pred. No. 2.5e-77;
 Matches 328; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

1 MAQPTPGGSLDWATVTEERAKHDQFLSKPIAGFITGDQARNFPGSGLPQVLAQIWMALADNMND 60


```

Db 1 MAQFPTPGSLDVAITVEERAKHDQGFHSLKPISTGDTGQANFFPGSLGPVLAQ 60
Qy 61 IVALADNNNDGMDQVEFSIAMKLIKLOGLPSTLPVWKKQPVAISSAPAFIGICI 120
Db 61 IVALADNNNDGMDQVEFSIAMKLIKLOGLPSTLPVWKKQPVAISSAPAFIGICI 120
Qy 121 ASMPPLTVAAPVPMGSIPIVGVSPVSSVPPAAPPVLANGAPVITGCLPAFAFAATWP 180
Db 121 ASMPPLTVAAPVPMGSIPIVGVSPVSSVPPAAPPVLANGAPVITGCLPAFAFAATWP 180
Qy 181 KSSFSRSGSGSLQNTKLQKQSPDVASAPPAEAVPQSSRLKRYROLFNHSDKTMGHL 240
Db 181 KSSFSRSGSGSLQNTKLQKQSPDVASAPPAEAVPQSSRLKRYROLFNHSDKTMGHL 240
Qy 241 TGPQARTILMOSSLTQOQLASIMNLSIDIDQDKLTAEFFILAMHLIDVAMSGQLPVL 300
Db 241 TGPQARTILMOSSLTQOQLASIMNLSIDIDQDKLTAEFFILAMHLIDVAMSGQLPVL 300
Qy 301 PEYIPFRRVRSGSGSVISSVSDQRLPEEPPSEDEQOPEKK 344
Db 301 PEYIPFRRVRSGSGSVISSVSDQRLPEEPPSEDEQOPEKK 344

```

RESULT 8

```

ID 081W15 PRELIMINARY; PRT; 611 AA.
AC 081W15;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DE 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Similar to interectin 2 (Fragment).
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC Tissue=Skin;
RA Struhsberg R.;
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC038963; AAH38963.1; -.
FT NON TER 611
SQ SEQUENCE 611 AA; 69236 MW; 85F6EA0C664A881A CRC64;

```

Query Match 24.4%; Score 1330; DB 4; Length 611;
 Best Local Similarity 54.0%; Pred. No. 2.8e-69;
 Matches 331; Conservative 86; Mismatches 118; Indels 78; Gaps 14;

```

Qy 1 MAQFPTPGSLDVAITVEERAKHDQGFHSLKPISTGDTGQANFFPGSLGPVLAQ 60
Db 1 MAQFPTPGSLDVAITVEERAKHDQGFHSLKPISTGDTGQANFFPGSLGPVLAQ 60
Qy 61 IVALADNNNDGMDQVEFSIAMKLIKLOGLPSTLPVWKKQPVAISSAPAFIGICI 118
Db 61 IVALADNNNDGMDQVEFSIAMKLIKLOGLPSTLPVWKKQPVAISSAPAFIGICI 118
Qy 119 GIASMPPLTVAAPVPMGSIPIVGVSPVSSVPPAAPPVLANGAPVITGCLPAFAFAATWP 165
Db 119 GIASMPPLTVAAPVPMGSIPIVGVSPVSSVPPAAPPVLANGAPVITGCLPAFAFAATWP 165
Qy 166 IOPPLFAFAFAATWPVPMGSIPIVGVSPVSSVPPAAPPVLANGAPVITGCLPAFAFAATWP 212
Db 166 IOPPLFAFAFAATWPVPMGSIPIVGVSPVSSVPPAAPPVLANGAPVITGCLPAFAFAATWP 212
Qy 212 IOPPLFAFAFAATWPVPMGSIPIVGVSPVSSVPPAAPPVLANGAPVITGCLPAFAFAATWP 229
Db 212 IOPPLFAFAFAATWPVPMGSIPIVGVSPVSSVPPAAPPVLANGAPVITGCLPAFAFAATWP 229
Qy 229 IOPPLFAFAFAATWPVPMGSIPIVGVSPVSSVPPAAPPVLANGAPVITGCLPAFAFAATWP 242
Db 229 IOPPLFAFAFAATWPVPMGSIPIVGVSPVSSVPPAAPPVLANGAPVITGCLPAFAFAATWP 242
Qy 242 IOPPLFAFAFAATWPVPMGSIPIVGVSPVSSVPPAAPPVLANGAPVITGCLPAFAFAATWP 289
Db 242 IOPPLFAFAFAATWPVPMGSIPIVGVSPVSSVPPAAPPVLANGAPVITGCLPAFAFAATWP 289
Qy 289 IOPPLFAFAFAATWPVPMGSIPIVGVSPVSSVPPAAPPVLANGAPVITGCLPAFAFAATWP 302
Db 289 IOPPLFAFAFAATWPVPMGSIPIVGVSPVSSVPPAAPPVLANGAPVITGCLPAFAFAATWP 302
Qy 302 IOPPLFAFAFAATWPVPMGSIPIVGVSPVSSVPPAAPPVLANGAPVITGCLPAFAFAATWP 349
Db 302 IOPPLFAFAFAATWPVPMGSIPIVGVSPVSSVPPAAPPVLANGAPVITGCLPAFAFAATWP 349
Qy 349 IOPPLFAFAFAATWPVPMGSIPIVGVSPVSSVPPAAPPVLANGAPVITGCLPAFAFAATWP 362
Db 349 IOPPLFAFAFAATWPVPMGSIPIVGVSPVSSVPPAAPPVLANGAPVITGCLPAFAFAATWP 362

```

```

Db 350 LVPSFR-----GKQI---DSINGTLPSTYOKMG-ESEPQKLPVTEDKRYANERGM 400
Qy 363 ELEKRRQALBEOQKEOEORLALQERAEQERERQOEAKROLLEKOLEKOREROR 422
Db 401 ELEKRRQALBEOQKEOEORLALQERAEQERERQOEAKROLLEKOLEKOREROR 460
Qy 423 EERREKTEREAKRELEROROLEMERNRROELLNORNKROEGTVLKARKTLEFELE 482
Db 461 EERREKTEREAKRELEROROLEMERNRROELLNORNKROEGTVLKARKTLEFELE 520
Qy 483 ALNDKHLQELQDIDICRLATQROEIESTNKRSLRIATITLQOQLOESQOMGLRILP 542
Db 521 ALNDKHLQELQDIDICRLATQROEIESTNKRSLRIATITLQOQLOESQOMGLRILP 580
Qy 543 EKQILSDQKQVQ 555
Db 581 EKQILSDQKQVQ 593

```

RESULT 9

```

ID 09U2T9 PRELIMINARY; PRT; 1097 AA.
AC 09U2T9;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Y116A8C.36 protein.
GN Y116A8C.36.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
OC Rhabditidae; Peloderiinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA McMurray A.A.;
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C. elegans: A platform for
RT investigating biology.";
RT Science 282:2012-2018 (1998).
CC -!- SIMILARITY: CONTRAINS 5 SH3 DOMAINS.
DR EMBL; AL117204; CAB55138.1; -.
DR HSSP; P29355; 1SEM.
DR WormPep; Y116A8C.36; CE23342.
DR InterPro; IPR000194; ATPase a/bcentre.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR000261; ERS15_homology.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00036; ehand; 2.
DR PRINTS; PR00452; SH3DOMAIN.
DR ProDom; PD000666; SH3; 5.
DR SMART; SM00027; EH; 2.
DR SMART; SM00326; SH3; 5.
DR PROSITE; PS00152; ATPASE_ALPHA_BETA; 1.
DR PROSITE; PS00018; EF_HAND; 2.
DR PROSITE; PS50031; EH; 2.
DR PROSITE; PS50002; SH3; 5.
KW SH3 domain.
SQ SEQUENCE 1097 AA; 122072 MW; 1C2BA5F103968372 CRC64;

```

Query Match 21.3%; Score 1333; DB 5; Length 1097;
 Best Local Similarity 29.4%; Pred. No. 4.8e-59;
 Matches 369; Conservative 200; Mismatches 454; Indels 234; Gaps 39;

```

Qy 15 WAITVEERAKHDQGFHSLKPISTGDTGQANFFPGSLGPVLAQIIVALADM 67
Db 5 WEVSDAEYQKFAF-----GQLTGQGPMDAVTARNALMRNLPTQVLSTQIIVALSDL 57

```

QY	68	NNDDGMDQVSEI	IAKKLILKLIQOQVLESTLIPVWKKQPVAISSAPAGIGIAEMPLY	127
Dd	58	DDQGLDIREIS	IAHMLALNCLAGIPTRPQLESSLVWA--RNAPRTWPGS-----	107
QY	128	AVAPRMGSI	PVWGMSPLV-----SSVPRAVPLNAGAPVIOPLPAFHPATWPK	181
Dd	108	-----RHGSVDV	SOQLPFAIDRMGSYIIPSA--FVSIAGTP-----	142
QY	182	SSSFERSG	BGSOQLTKLOKASFDVASAPRAEWAVPOSSRLKYROLFNSHDKTNSGHLT	241
Dd	143	SSRHSIS	SAGSELYN-----NDRNVEGRQLENNAIIPHNNKLKYQLFNALDKETLGISLS	196
QY	242	GPOARTIL	MQSSLPOAQASIMWNLSDDODGLTMEFLLMHLIDVAMSOQLPVPVLP	301
Dd	197	SOVGSALG	LSGLPFTNLVAHIWFLSDVNKDLGIDEVAYISQYIMEMKSGALPKIRPL	256
QY	302	EYIPSPFR	VSGSGVSISSSSVQRLPEEPSESDCOPEKULVTFEDDKREKPEEGS	361
Dd	257	ELV-----	-----RMGSISSRSANNTPELEGAPPO--KSPAKTIEDKQDNLSSKQ	303
QY	362	VELEKRO	ALLEQOREQERLAOLERAQERKERQEOZAKROLELEKOLEKORELRO	421
Dd	304	AELERRR	RLVEEBQRRRAVEKKEEBEAKNRRQERQERQDAVERQALELROQILIAQ	363
QY	422	REEE----	RRKEIERR-----BAKKELE---ROQOLEMERNROELLNORNKQEGTVV	469
Dd	364	REEEKKR	RLLEMERREDEDEKKRKQOMEKAKVQOVPEQKKNFYQOKOEENRLAQ	423
QY	470	LKARKTE	LEFELALNDKKHOLEGLDIDRCGLATQROEIESNTKSRLELAIETHLOOQ	529
Dd	424	ROQRKTI	LOFOLALDEKVIDEVIDIGAKAENAVETGTEIMSTRQEKVARI---KE	479
QY	530	LOESQOM	GRLLPEKOLISDLQOVQONSLSHRSDLLTKRALEAKELARQOLEQD---	586
Dd	480	LOETNQ---	KTAIESQELGHOL--LOQSAHMET--TQKSELLEALRRRBDARIMAIIDAA	533
QY	587	-EVERETS	SKLOEIVF--NOLUKEIREHISQOLOKORSLEAARLIKQYQOEKRSILEKQ	644
Dd	534	LELSTEK	ESYQUTILKTNKEKYUTDYSK-----LVAKKEEYNSFEL---	578
QY	645	KEDARV	ROERNKQMLEHVQOEORPRKP-----HEEDRLKRDYSRK	688
Dd	579	-----LVNA	OTNARSKIGEFKASAPASAPAPAPARTTNGFPANNDNAPGEFD--K	631
QY	689	KEAERAP	REMOKOSRLFHPHOEPAKLATQAPWSTTEKGPLTISAOESVAKVYRALYR	748
Dd	632	TDASGR	PADFEATST-----ADFPALQAP--AHKGAVDQSAFNIHDTYKRALFA	683
QY	749	FESRSH	DEITTOGDIVMDSQTEPCWLGELGKGTGWPRANYAEKIPENEVPTPRKP	808
Dd	684	FEARSE	DELSEFPBGVITVIFQSHAAPEPMRAGOLEKYGWPEAPEVAIA--AVPTPG--	739
QY	809	VTDLSA	PAPKALAFETAPRLPVYSESESTTPNMNADSSITWPSSNKEKPEFDNMDTAA	868
Dd	740	-----GDPR	IQMPPMPTWSSSVDDIGV---KAARKAEI-----AA	772
QY	869	QPSLTV	PAGOLRQSAFTPATGSSSPVLQGEKEVGLQAOALYPMRAKDNHNLFPN	928
Dd	773	AMGLTEGA	-----PRASSAPAAANVIS-----QCLAQOFQMRANEDDLSFA	814
QY	929	KSDVITV	LEQODMMFG--EVQOGKQWFPKSYVKLISGVRKSTSIDTGPTEPSASLKRA	987
Dd	815	KGDTLEV	ELEKQMKMKGNHPAGEIQWFPKSYKVEGATTSPTTPI--VSPKSAQAPGA	873
QY	988	SPAAPAL	PGBE-----FIAMTYESESQDLPFOQGDVIVTTKQGDWMTG	1034
Dd	874	AGAQYDV	VPDVTLOASETAPQOOLYTVIYPEAVETTDLALHVDITLIVLEKDNEMWK	933
QY	1035	TVGDGSG	VFPNSVYVRLKDSGEGTAGTKV--GSLGKKPEI---AQVIASVYATGPOLTLA	1089
Dd	934	RONGREG	IFPANYVVISVOQAGDPPRPQAPRPAAPRVLYCEAKVVDVFAVASAPOLGIX	993
QY	1090	PGQLIL	KKKQNGWMBEGLQARKKROIGWEPANYVULSLPGTSKITPTBLPKTAVQPA	1149

Db	Query Match	Best Local Similarity	Matches	21.0%; Score 1314; DB 11; Length 755; Pred. No. 2.7e-58; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Db	994 VGEIVTKIREKSAAGMWEGLIRNGKP-IAGWPFEEYVKVILEAASPAT-----	1040		
Qy	1150 VCOYIAGMDYTAQNDDLEAFSGQGIINVINKEDPDMWKG---EWSGGVGLFPSPNYVK	1203		
Db	1041 --RATAYVDYEAASGPDELGFRTGVIVITVTKSEAFMWSGIREQDPSSGGLFPSPNYVQ	1095		
RESULT 10				
Q8C4B5				
ID	PRELIMINARY;	PRT;	755 AA.	
AC	Q8C4B5			
DT	01-MAR-2003 (TREMBLrel. 23, Created)			
DT	01-MAR-2003 (TREMBLrel. 23, Last sequence update)			
DT	01-MAR-2003 (TREMBLrel. 23, Last annotation update)			
DE	Intersectin (Fragment).			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognath; Muridae; Murinae; Mus.			
NCBI_TextID=10090;				
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6J; TISSUE=Cerebellum;			
RX	MEDLINE=22354683; PubMed=12466851;			
RA	The FANTOM Consortium,			
RT	the RIKEN Genome Exploration Research Group Phase I & II Team;			
RT	"Analysis of the mouse transcriptome based on functional annotation of			
RU	60,770 full-length cDNAs."			
DR	Nature 420:563-573(2002).			
EMBL	AK082606; BAC38546.1; -.			
NON_TER	1			
FT	SEQUENCE 755 AA; 85107 MW; 8FD0026865297597 CRC64;			
Qy	964 GPRKSTSIDGPRESASLKRVASPAKPAIPGEERFIAMTYESSSQGLTFQGGVIV	1023		
Db	5 GPRKSTSIDGPRESASLKRVASPAKPAIPGEERFIAMTYESSSQGLTFQGGVIV	64		
Qy	1024 VTKKDGWMTGTVDKSGVFPSPNYVRDKDSGSGTAGTGS LGKPEIAVIVASYAATGP	1083		
Db	65 VTKKDGWMTGTVDKSGVFPSPNYVRDKDSGSGTAGTGS LGKPEIAVIVASYAATGP	124		
Qy	1084 EQLTLAAGQILIRKKNPGMWESEELQARKKQIQGFPPANVYKLSPGSKITPEELPK	1143		
Db	125 EQLTLAAGQILIRKKNPGMWESEELQARKKQIQGFPPANVYKLSPGSKITPEELPK	184		
Qy	1144 TAVOPAVCOVIGMDYTAQNDDLEAFSGQGIINVINKEDPDMWKGEVSGGVGLFPSPNYVK	1203		
Db	185 TAVOPAVCOVIGMDYTAQNDDLEAFSGQGIINVINKEDPDMWKGEVSGGVGLFPSPNYVK	244		
Qy	1204 LTTMDPSQQ 1213			
Db	245 LTTMDPSQQ 254			
RESULT 11				
Q8C9C3				
ID	PRELIMINARY;	PRT;	440 AA.	
AC	Q8C9C3			
DT	01-MAR-2003 (TREMBLrel. 23, Created)			
DT	01-MAR-2003 (TREMBLrel. 23, Last sequence update)			
DT	01-MAR-2003 (TREMBLrel. 23, Last annotation update)			
DE	SH3 domain protein 1B (Fragment).			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognath; Muridae; Murinae; Mus.			
NCBI_TextID=10090;				
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6J; TISSUE=Thymus;			

RX MEDLINE=22354683; PubMed=12466851;
 RA The RANTOM Consortium;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs."
 RL Nature 420:563-573(2002).
 DR EMBL; AK044449; BAC31264.1; -
 FT NON TER 440
 SQ SEQUENCE 440 AA; 48760 MW; 1C94A5E62B71EEF2 CRC64;

Query Match 18.7%; Score 1172; DB 11; Length 440;
 Best Local Similarity 56.7%; Pred. No. 2e-51;
 Matches 261; Conservative 58; Mismatches 87; Indels 54; Gaps 15;

QY 1 MAQPTPGGSLDVAITVEERAKDQFLSKPIAGITGDQANFFFGSLPPVLAQ 60
 DB 1 MAQPTPGGSLDVAITVEERAKDQFLSKPIAGITGDQANFFFGSLPPVLAQ 60
 QY 61 IVALADNMNDGRMDQVERSIAMKLIKLOGLPSTLPVWKQOPV--ATSSAPAFGIG 118
 DB 61 IVALADNMNDGRMDQVERSIAMKLIKLOGLPSTLPVWKQOPV--ATSSAPAFGIG 118
 QY 119 GIASNP-----PLTAVAPV--PMG-----SIPVAGSPPLVSSVPPAAVPLANGAP 164
 DB 120 ---SWPNLSIHQPLPVPVPIATPLSSAGTSTPLPLMAPLPVPSVTSLSLP--NGTAS 173
 QY 165 VIQPLPAHPAATPKSSFS--RSGPGSGLNTKLOKAQ--FDVASAP----- 211
 DB 174 LIQPL-SIPYSSSTLPHASYSYSLMNGCFG--GASIQKASLIDIGSSSSSTASLSGN 229
 QY 212 -----AAEVAPOSRLKYROLFNSHDQTMGSLGPOARTIMOSLPQOALASIMLS 266
 DB 220 SPKGTSEWAVPQPRKLRQKFNSLDKMSGYLSGFORAKNALDSNLSQTLATITWLA 289
 QY 267 DIDQDKLTAEFFILAMHLIDVAMSGQLPVLPEPIYIPSEFRVSSGMSVSSSD 326
 DB 290 DIDGQGLKAEFFILAMHLIDVAMSGQLPVLPEPIYIPSEFRVSSGMSVSSSD 341
 QY 327 QRLPEPSSSEDEQOEKLPVTFEKKRENFRRGVELEKROALLEQKQERLAQLE 386
 DB 342 GTLPYQQTQ--BEERQKLPVTFEKKRENFRRGVELEKROALLEQKQERLAQLE 400
 QY 367 RAEQKEREEROEAKROLLEKOLEKORELEKOREER 426
 DB 401 KEWEEKORELOEQEMKQLEKLEKRELEKORELEKOREER 440

RESULT 12

08CJ43 PRELIMINARY; PRT; 164 AA.

AC 08CJ43; 08CJ43;
 DT 01-MAR-2003 (Tremblrel. 23, Created)
 DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
 DE Intersectin isoform 4 (Fragment).
 GN ITSN.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NX NCBI_TaxId=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BAIB/cx57BLP1; TISSUE=lung;
 RA Skripkina I.Ya., Teyba L.A., Slavov D., Gardiner K., Rynditch A.V.;
 RT "Novel splicing forms of human and mouse intersectin gene
 RT transcripta."
 RL Ukr. Biokhim. Zh. 74:33-43(2002).
 DR EMBL; AF525079; AAN75695.1; -
 FT NON TER 1
 FT NON TER 164
 SQ SEQUENCE 164 AA; 17677 MW; 5E5868CA919DE969 CRC64;

Query Match 13.1%; Score 822.5; DB 11; Length 164;

Best Local Similarity 81.6%; Pred. No. 2.3e-34;
 Matches 164; Conservative 0; Mismatches 0; Indels 37; Gaps 1;

QY 2 AQPPTPGGSLDVAITVEERAKDQFLSKPIAGITGDQANFFFGSLPPVLAQ 61
 DB 1 AQPPTPGGSLDVAITVEERAKDQFLSKPIAGITGDQANFFFGSLPPVLAQ 60
 QY 62 WALADNMNDGRMDQVERSIAMKLIKLOGLPSTLPVWKQOPV--ATSSAPAFGIG 121
 DB 61 WALADNMNDGRMDQVERSIAMKLIKLOGLPSTLPVWKQOPV--ATSSAPAFGIG 114
 QY 122 SMPPLTAVAPVPMGSIIPVAGSPPLVSSVPPAAVPLANGAPVLIQPLPAFPAATWPK 181
 DB 115 -----AAVPLANGAPVLIQPLPAFPAATWPK 143
 QY 182 SSSFSRSGPGSGLNTKLOKAQ 202
 DB 144 SSSFSRSGPGSGLNTKLOKAQ 164

RESULT 13

08CGU5 PRELIMINARY; PRT; 158 AA.

AC 08CGU5;
 DT 01-MAR-2003 (Tremblrel. 23, Created)
 DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
 DE Intersectin isoform 5 (Fragment).
 GN ITSN.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NX NCBI_TaxId=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BAIB/cx57BLP1; TISSUE=Brain;
 RA Skripkina I.Ya., Teyba L.A., Slavov D., Gardiner K.;
 RT "Novel splicing forms of human and mouse intersectin gene
 RT transcripta."
 RL Ukr. Biokhim. Zh. 74:33-43(2002).
 DR EMBL; AY127576; AAN02285.1; -
 FT NON TER 1
 FT NON TER 158
 SQ SEQUENCE 158 AA; 17252 MW; 4DD3D8D070CC4E5 CRC64;

Query Match 13.0%; Score 812.5; DB 11; Length 158;
 Best Local Similarity 96.8%; Pred. No. 7.1e-34;
 Matches 153; Conservative 0; Mismatches 0; Indels 5; Gaps 1;

QY 708 HHQEPKLTQAQWSTTEKGLTISAQESKVVYVYALVPFESRSHDEITIQGDIVM- 766
 DB 1 HHQEPKLTQAQWSTTEKGLTISAQESKVVYVYALVPFESRSHDEITIQGDIVM 60
 QY 767 -----VDESQGEPMGLGELKGTGWFPANYAEKIPENEVPTPAKPYTDLTSAPAKLAL 822
 DB 61 KGEWVDESQGEPMGLGELKGTGWFPANYAEKIPENEVPTPAKPYTDLTSAPAKLAL 120
 QY 823 RETPAPLPVTISSEPTTPNNWADFSSTWPSSSNEKPEPT 860
 DB 121 RETPAPLPVTISSEPTTPNNWADFSSTWPSSSNEKPEPT 158

RESULT 14

08T068 PRELIMINARY; PRT; 662 AA.

AC 08T068;
 DT 01-JUN-2002 (Tremblrel. 21, Created)
 DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
 DE LD23686P.
 GN DAPI60 OR CG1099.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 4, 2003, 21:34:26 ; Search time 770.054 Seconds
(without alignments)
15719.093 Million cell updates/sec

Title: US-09-674-237a-2

Perfect score: 3642
Sequence: 1 atggctcagcttcacccacc.....tgaccacagccagcaatga 3642

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 2201672 seqs, 1661799599 residues

Total number of hits satisfying chosen parameters: 4403344

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first: 45 summaries

Database: Published Applications NA:

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	477.4	13.1	2017	10	US-09-884-441-72
2	477.4	13.1	2017	11	US-09-907-969-72
3	477.4	13.1	2017	12	US-09-827-271-72
4	477.4	13.1	2017	14	US-10-198-053-72
5	356	9.8	568	11	US-09-764-881-55
6	332.4	9.1	2873	9	US-09-879-957-193
7	331.4	9.1	4210	10	US-09-764-868-125
8	322.6	8.9	503	11	US-09-918-995-31258
9	258	7.1	747	9	US-09-879-957-39
10	198.8	5.5	270	9	US-09-864-761-17127
11	198.8	5.5	286	9	US-09-864-761-26348
12	198.8	5.5	297	9	US-09-864-761-30453
13	198.6	5.5	301	9	US-09-864-761-17146
14	197.4	5.4	263	9	US-09-864-761-17644
15	192	5.3	480	9	US-09-864-761-10314
16	180.8	5.0	486	9	US-09-864-761-333

17	180.8	5.0	487	9	US-09-864-761-864	Sequence 864, App
18	163	4.5	475	9	US-09-864-761-311	Sequence 311, App
19	163	4.5	475	9	US-09-864-761-13884	Sequence 13884, A
20	154.6	4.2	967	11	US-09-764-881-50	Sequence 50, App1
21	146.4	4.0	180	9	US-09-864-761-17125	Sequence 17125, A
22	146.2	4.0	292	9	US-09-864-761-20261	Sequence 20261, A
23	144.2	4.0	304	9	US-09-864-761-21373	Sequence 21373, A
24	144.2	4.0	310	9	US-09-864-761-19751	Sequence 19751, A
25	144.2	4.0	310	9	US-09-864-761-19759	Sequence 19759, A
26	144	4.0	480	10	US-09-884-441-60	Sequence 60, App1
27	144	4.0	480	11	US-09-907-969-60	Sequence 60, App1
28	144	4.0	480	12	US-09-827-271-60	Sequence 60, App1
29	144	4.0	480	14	US-10-198-053-60	Sequence 60, App1
30	144	4.0	531	10	US-09-884-441-5	Sequence 5, App1
31	144	4.0	531	11	US-09-907-969-5	Sequence 5, App1
32	144	4.0	531	12	US-09-827-271-5	Sequence 5, App1
33	144	4.0	531	14	US-10-198-053-5	Sequence 5, App1
34	144	4.0	3489	12	US-10-294-804-1	Sequence 1, App1
35	142.4	3.9	955	11	US-09-764-881-54	Sequence 54, App1
36	140.4	3.9	230	9	US-09-864-761-17643	Sequence 17643, A
37	140.4	3.9	247	9	US-09-864-761-30501	Sequence 30501, A
38	132.2	3.6	462	11	US-09-918-995-21728	Sequence 21728, A
39	127	3.5	419	9	US-09-864-761-309	Sequence 309, App
40	125.2	3.4	484	9	US-09-864-761-863	Sequence 863, App
41	121.6	3.3	531	10	US-09-884-441-4	Sequence 4, App1
42	121.6	3.3	531	11	US-09-907-969-4	Sequence 4, App1
43	121.6	3.3	531	12	US-09-827-271-4	Sequence 4, App1
44	121.6	3.3	531	14	US-10-198-053-4	Sequence 4, App1
45	119.2	3.3	1926	12	US-10-294-804-3	Sequence 3, App1

ALIGNMENTS

RESULT 1	US-09-884-441-72	Application US/09884441
Sequence 72, App1	Parent No. US20020119158A1	
GENERAL INFORMATION:		
APPLICANT:	Algate, Paul A.	
TITLE OF INVENTION:	COMPOSITIONS AND METHODS FOR THE THERAPY AND	
FILE REFERENCE:	210121.462C7	
CURRENT APPLICATION NUMBER:	US/09/884,441	
CURRENT FILING DATE:	2001-06-18	
NUMBER OF SEQ ID NOS:	489	
SOFTWARE:	FaastSeq for Windows Version 3.0	
SEQ ID NO 72		
LENGTH:	2017	
TYPE:	DNA	
ORGANISM:	Homo sapien	
US-09-884-441-72		
Query Match	13.1%;	Score 477.4; DB 10; Length 2017;
Best Local Similarity	55.5%;	Pred. No. 8.6e-127;
Matches 111;	Conservative 0;	Mismatches 796; Indels 96; Gaps 6;
1	ATGGCTCAGTTTCCACACCTTTCGTTGCTGATGCTGGGCCATCTGTGAG	60
2	ATGGCTCAGTTTCCACACCTTTCGTTGCTGATGCTGGGCCATCTGTGAG	60
3	ATGGCTCAGTTTCCACACCTTTCGTTGCTGATGCTGGGCCATCTGTGAG	60
4	ATGGCTCAGTTTCCACACCTTTCGTTGCTGATGCTGGGCCATCTGTGAG	60
5	ATGGCTCAGTTTCCACACCTTTCGTTGCTGATGCTGGGCCATCTGTGAG	60
6	ATGGCTCAGTTTCCACACCTTTCGTTGCTGATGCTGGGCCATCTGTGAG	60
7	ATGGCTCAGTTTCCACACCTTTCGTTGCTGATGCTGGGCCATCTGTGAG	60
8	ATGGCTCAGTTTCCACACCTTTCGTTGCTGATGCTGGGCCATCTGTGAG	60
9	ATGGCTCAGTTTCCACACCTTTCGTTGCTGATGCTGGGCCATCTGTGAG	60
10	ATGGCTCAGTTTCCACACCTTTCGTTGCTGATGCTGGGCCATCTGTGAG	60
11	ATGGCTCAGTTTCCACACCTTTCGTTGCTGATGCTGGGCCATCTGTGAG	60
12	ATGGCTCAGTTTCCACACCTTTCGTTGCTGATGCTGGGCCATCTGTGAG	60
13	ATGGCTCAGTTTCCACACCTTTCGTTGCTGATGCTGGGCCATCTGTGAG	60
14	ATGGCTCAGTTTCCACACCTTTCGTTGCTGATGCTGGGCCATCTGTGAG	60
15	ATGGCTCAGTTTCCACACCTTTCGTTGCTGATGCTGGGCCATCTGTGAG	60
16	ATGGCTCAGTTTCCACACCTTTCGTTGCTGATGCTGGGCCATCTGTGAG	60
17	ATGGCTCAGTTTCCACACCTTTCGTTGCTGATGCTGGGCCATCTGTGAG	60
18	ATGGCTCAGTTTCCACACCTTTCGTTGCTGATGCTGGGCCATCTGTGAG	60
19	ATGGCTCAGTTTCCACACCTTTCGTTGCTGATGCTGGGCCATCTGTGAG	60
20	ATGGCTCAGTTTCCACACCTTTCGTTGCTGATGCTGGGCCATCTGTGAG	60
21	ATGGCTCAGTTTCCACACCTTTCGTTGCTGATGCTGGGCCATCTGTGAG	60
22	ATGGCTCAGTTTCCACACCTTTCGTTGCTGATGCTGGGCCATCTGTGAG	60
23	ATGGCTCAGTTTCCACACCTTTCGTTGCTGATGCTGGGCCATCTGTGAG	60
24	ATGGCTCAGTTTCCACACCTTTCGTTGCTGATGCTGGGCCATCTGTGAG	60
25	ATGGCTCAGTTTCCACACCTTTCGTTGCTGATGCTGGGCCATCTGTGAG	60
26	ATGGCTCAGTTTCCACACCTTTCGTTGCTGATGCTGGGCCATCTGTGAG	60
27	ATGGCTCAGTTTCCACACCTTTCGTTGCTGATGCTGGGCCATCTGTGAG	60
28	ATGGCTCAGTTTCCACACCTTTCGTTGCTGATGCTGGGCCATCTGTGAG	60
29	ATGGCTCAGTTTCCACACCTTTCGTTGCTGATGCTGGGCCATCTGTGAG	60
30	ATGGCTCAGTTTCCACACCTTTCGTTGCTGATGCTGGGCCATCTGTGAG	60
31	ATGGCTCAGTTTCCACACCTTTCGTTGCTGATGCTGGGCCATCTGTGAG	60
32	ATGGCTCAGTTTCCACACCTTTCGTTGCTGATGCTGGGCCATCTGTGAG	60
33	ATGGCTCAGTTTCCACACCTTTCGTTGCTGATGCTGGGCCATCTGTGAG	60
34	ATGGCTCAGTTTCCACACCTTTCGTTGCTGATGCTGGGCCATCTGTGAG	60
35	ATGGCTCAGTTTCCACACCTTTCGTTGCTGATGCTGGGCCATCTGTGAG	60
36	ATGGCTCAGTTTCCACACCTTTCGTTGCTGATGCTGGGCCATCTGTGAG	60
37	ATGGCTCAGTTTCCACACCTTTCGTTGCTGATGCTGGGCCATCTGTGAG	60
38	ATGGCTCAGTTTCCACACCTTTCGTTGCTGATGCTGGGCCATCTGTGAG	60
39	ATGGCTCAGTTTCCACACCTTTCGTTGCTGATGCTGGGCCATCTGTGAG	60
40	ATGGCTCAGTTTCCACACCTTTCGTTGCTGATGCTGGGCCATCTGTGAG	60
41	ATGGCTCAGTTTCCACACCTTTCGTTGCTGATGCTGGGCCATCTGTGAG	60
42	ATGGCTCAGTTTCCACACCTTTCGTTGCTGATGCTGGGCCATCTGTGAG	60
43	ATGGCTCAGTTTCCACACCTTTCGTTGCTGATGCTGGGCCATCTGTGAG	60
44	ATGGCTCAGTTTCCACACCTTTCGTTGCTGATGCTGGGCCATCTGTGAG	60
45	ATGGCTCAGTTTCCACACCTTTCGTTGCTGATGCTGGGCCATCTGTGAG	60

241 GCCATGAGCTTATGAACTGAGTACAGAGATATGAGTCCCTCCACTTCCCT 300
 Db GCTATGAACTCATCATGTTAAAGTTGAGGCGCAAGCTGCTGATGCTCTCTCT 341
 Qy 301 GTCATGAAACGCAACCACTG---GCTATTTCCAGTGCACAGCATTTGTATAGAGG 357
 Db 342 ATCATGAAACCAACCCCTATGTTCTCTCACTAATCTCTGCTCGTTTGGATGGAAAG 401
 Qy 358 ATGCTAGCATG-----CCACCACTCAAGCTGTTGCTCTGTCCTCA----- 400
 Db 402 ATGCCCATCTGTCATTCAATGAGCATTTGCTCAGTTGCACTATGCAAAACCTTGG 461
 Qy 401 -----TGGCTCCATTGCAAGTTGGAATGTCTCCACCTTATGATCT 444
 Db 462 TCTTCTGCTACTTCAAGGACCAAGTATCTCTCCCTAATGATGCTGCTCTCTGCTGCT 521
 Qy 445 TCTGTCCTCCAGCAGAGTGCCTCCCTGCTAAGGGGCTCTCCCTGATACAGCT 504
 Db 522 TCTGTTAGTACATCTCATTAACCAATGGAAGTGCATCTCATTTCAAGCTTTATCAT 581
 Qy 505 CTGCTGCGTTTGGCGCATCTGACGCAATGCGCAAGAGTTCTTCTTCAAGCATCT 564
 Db 582 CTTATTTCTTCTTCAACATTTGCTCATCATCTTACAGCTGATGATGGAGATTT 641
 Qy 555 GGTCCAGGCTCACTAATTAACCTAAGTTACAGAGGCAATCATTTGATGCTGCCAGC 624
 Db 642 GGTGTTGCTAGTATCAAGAGGCGCAATCTGATTTAGTTAGTTAGTACTCAACT 701
 Qy 625 GCGCTCCAG-----CAGCAAAATGGGCTGTG 651
 Db 702 TCTCTAATGCTCTCCCTCTCAGAGGAATCACCCTAAGACAGGAGCTCAGAGTGGCAGT 761
 Qy 652 CTTGATCATCAGGCTGAATATACAGCAATTTTCAACAGCAGCAACAACTATAGT 711
 Db 722 CTTGAGCTTCAAGATTAAGTAAAGTAAAGTAAATTAATAGCTAGACAAAGGCTAGAGC 821
 Qy 712 GGAACCTTAAGAGTCCCGAGCAGAACTATTTCAAGCATCAAGTTTACCCAGCT 771
 Db 822 GGAATCCTCTCAGGTTTCAAGCTAAGAAATGCTTTTCAAGTCAAACTCTCTCAACT 881
 Qy 772 CAGTGGCTTCAATATGAACTTTCTGACATTTGATCAAGATGAAATCACTGACAGAA 831
 Db 882 CAGTACGCTACTATTTGAGCTCTGCTGATCATGATGATGAGAGCAGCTTGAAGCTGAA 941
 Qy 832 GAATTTATCTAGCTATGACCTAATTAATGATGCTGCTGCTGCTGCTGCTGCTGCTG 891
 Db 942 GAATTTATCTGCGATGACCTCACTGACATGCGCAAGCTGGAAGCCTACATGCTG 1001
 Qy 892 GTTCTGCTCCAGATATCATCCCTCTCTTCAAGAAAGTTGCTGCGGAGCTGGAGT 951
 Db 1002 AGTTGCTCTCCAGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1050
 Qy 952 TCCGTCATAGCTCTTCTTCTGATGATCAGAGCTGCTGAGAGCCCTGCTCAGAGAT 1011
 Db 1051 -----TTGATTTCTTTAATGAACTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1097
 Qy 1012 GAGCAGCAGCAGAGAAAGAACTGCTGATTTGATTAAGATTAAGAGCGGAGAACTTC 1071
 Db 1098 GA---AGAGCCTCAAGAAAGAACTGCTGATTTGAGAGCAAAAGGAGCAACTAT 1154
 Qy 1072 GAGCAGAGAGTGTGAGAGTGTGAGAGAGCGCGCAGAGAGCTTGTGAGAGAGAGGCAAA 1131
 Db 1155 GAACAGAGAAACATGAGAGTGTGAGAGAGCGCAAGAGTGTGATGAGAGAGAGAGAGG 1214
 Qy 1132 GAGCAGAGAGGTTGCTCAGCTGAGAGCGCGCAGAGAGAGAGAGAGAGAGAGAGAGAG 1191
 Db 1215 GAGGCTGAAGCGAAAGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1274
 Qy 1192 CAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1251
 Db 1275 CAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1334

Qy 1252 CTGAGCGCGAGCGAGAGAGAGAGAGAGAGAGATCGAGAGCGCGAGAGAGAGAGAGAG 1311
 Db 1335 CTGAGAGAGAGAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1394
 Qy 1312 CCGGAACTGAAAGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1371
 Db 1395 CAGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1454
 Qy 1372 AATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1431
 Db 1455 AGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1514
 Qy 1432 GAGTTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1491
 Db 1515 CAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1574
 Qy 1492 ATCAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1551
 Db 1575 GTCCAAATCAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1634
 Qy 1552 CTAGAGATGCTGAAATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1611
 Db 1635 CTGAAATTTATGAAATCAAACTTCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1694
 Qy 1612 GGAAGATTTATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1671
 Db 1695 ATCTATCTGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1754
 Qy 1672 AGTTGATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1731
 Db 1755 AACACACTGATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1814
 Qy 1732 CCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1791
 Db 1815 TGCCAAAGAGCTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1874
 Qy 1792 GAGATGATGTTTCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1851
 Db 1875 GAATGATTCATTTAATCAATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1934
 Qy 1912 AGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1934
 Db 1995 AGATTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2017

RESULT 2
 US-09-907-969-72
 ; Sequence 72, Application US/0907969
 ; Publication No. US20030091580A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Micham, Jennifer L.
 ; APPLICANT: King, Gordon E.
 ; APPLICANT: Algate, Paul A.
 ; APPLICANT: Fling, Steven P.
 ; APPLICANT: Recter, Marc W.
 ; APPLICANT: Fanger, Gary Richard
 ; APPLICANT: Reed, Steven G.
 ; APPLICANT: Vedrick, Thomas S.
 ; APPLICANT: Carter, Patrick
 ; APPLICANT: Hill, Paul
 ; APPLICANT: Albone, Earl
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
 ; OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
 ; FILE REFERENCE: 210121.462C8
 ; CURRENT APPLICATION NUMBER: US/09/907,969
 ; NUMBER OF SEQ ID NOS: 596
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 72

LENGTH: 2017
TYPE: DNA
ORGANISM: Homo sapiens
US-09-907-969-72

Query Match 13.1% Score 477.4; DB 11; Length 2017;
Best Local Similarity 55.5% Pred. No. 8,6e-127;
Matches 1111; Conservative 0; Mismatches 796; Indels 96; Gaps 6;

QY 1 ATGGCTCAGTTTCCACACCTTTCGGTGTAGCTGTGATGTCTGGGCAATACCTGTGAG 60
DB 42 ATGGCTCAGTTTCCACACGATGAAATGAGGGCCAAATATGTGGCTATTATCAATCGAA 101
QY 61 GAAAGGCCCAGAGATGACAGAGTTCTTACGCTGAGCCGATAGCGGATTTATTAATCT 120
DB 102 GAAAGTACTAGATGATTAACAGTTTGATTAACCTTCAAGAGGTTACATTAACA 161
QY 121 GGTGATCAAGCGAAGAACTTTTTCATCTGGGTTACCTGAGCTGTCTTAGCACA 180
DB 162 GGTGATCAAGCCGTTCTTTTCTACAGTCAAGTCTGCGGCGCGGTTTATAGCTGAA 221
QY 181 ATATGGGCGCTAGCGACATGATTAAGATGAAAGATGATCAATGGAATTTTCATA 240
DB 222 ATATGGGCGCTTATCAATCTGAACAAGATGGAAGATGGAACGAGATTTCTCTATA 281
QY 241 GCCATGAAGCTTATCAAACTGAAGCTACAGAAATATGAGTCCCTCCACACTTCCCCCT 300
DB 282 GCATGAATCTATCAAGTTAAAGTTGACAGGCGCAAGCTGCTGTATGCTCTCTCT 341
QY 301 GTCATGAACAGAACCAAGT---GCTATTTCCAGTGCACAGCAATTTGGTATAGAGG 357
DB 342 ATCATGAACCAACCCCTATGTTCTCTCCACTATCTCTGCTGTTTGGATGGAGAGC 401
QY 358 ATATGCTAGATG-----CCACCACTCAAGCTGTGCTCTGTGCCA----- 400
DB 402 ATGCCCAATCTGTCATTCATTCAGCCATTCCTCCAGTTGCACTTATAGAACACCCCTTG 461
QY 401 -----TGGGCTCCATTTCCAGTTGTTGAATGTCACACCTTATGATCT 444
DB 462 TCTTCTGCTAATTCAGAGACAGATATTCCTCCCTTAATGATGCTGCTGCTGAGCT 521
QY 445 TCTGTCCTTCAGACAGAGTCCCTCCCTGCTTAAGGGGCTCTCCCTCATACAGCT 504
DB 522 TCTGTTAGTACATCTCATTAACAAATGAACTGCCAGTCTCATTCAGCTTATCAT 581
QY 505 CTGCTGCGTTTGGCATCTCTGACGCACTGGCCAAAGATTCTTCTTCAAGAGATCT 564
DB 582 CCTTATCTTCTTCAACATTTGCTCTCAATGCAATCTTACAGCTGATGATGGAGATTT 641
QY 565 GGTCCAGGGTCAACATTTAAACATAAGTTACAGAGGCAACATTCATTCATGTCGCGACG 624
DB 642 GGTGTGCTAGTATCCAGAAAGCCCAAGTCTCTGATTTAGATCTGATAGCTCAACT 701
QY 625 GCCCTTCAG-----CAGCAGATGGGCTGTG 651
DB 702 TCTTCAATGCTTCCCTTCAGGGAATCACTAAGACAGGACCTCAAGTGGGAGTT 761
QY 652 CCTCAATCAAGGCTGAAATACAGGAGTTTCAACAGCCACACAACTATGAGT 711
DB 762 CCTCAAGCTTCAAGATTTAAAGTATCGCAAAATTTAATGTACAAAGGATGAGC 821
QY 712 GGACATTTAAAGGTTCCCGACAGCAAGATCTCTATGCAATCAAGTTTATCCCGAGCT 771
DB 822 GGATACCTCTCAGGTTTTCAGCTAATAATGCCCTTCTTCAAGTCAATCTCTTCAACT 881
QY 772 CAGCTGCTTCAATATGATCTTTCTGACATTTGATCAAGATGGAATACTCACTGAGAA 831
DB 882 CAGCTGCTACTATTTGAGCTGTGGCTGACATGATGATGAGGAGAGATTTGAAGCTGAA 941
QY 832 GAATTTATCTTACGCTATGACCTTAATTTGATTTGCAATGTTGCTGAGGCACTGCGGCC 891
DB 942 GAATTTATCTTGGCGATGACCTCACTGACATGGCCAAAGCTGAGCAGGCACTACCACTG 1001

QY 892 GTCTGCTCCAGAAATACATCCCTCTCTTCCAGAAAGTTGCTCCGCGACATGGGATG 951
DB 1002 AGCTTGCTTCCGAGAGTTGCTTCTTCAATCTTTCAGAGGGGAGAAAGCAAG----- 1050
QY 952 TCCGTATTAAGCTCTTCTTCTGTGATCAGAGGCTGCTGAGAGGCGTGTGAGAGAT 1011
DB 1051 -----TTGATTTCTGTTATGGAACCTCTGCTTCAATATCAGAAAAACAAAGAA 1097
QY 1012 GAGCAGACCCAGAGAAAGAACTGCTGTGACATTTGAAAGATTAAGAGCGGGAATCTTC 1071
DB 1098 GA---AGACCTTCAAGAAAGAACTGCGATTAATTTTGAAGACAAAGGAAAGCCACTAT 1154
QY 1072 GAGCGAGGAGTGTGAGGCTGAGAAAGCCCGCAAGCCCTCTTGGAGAGCAGCGCAAA 1131
DB 1155 GAAAGGAGAAACATGAGAGCTGAGAAAGCAAGCCAAATTTGATGAGAGCAGCAGAGG 1214
QY 1132 GAGCAGAGCGGTTGCTCAGCTGAGCGCGCCGACAGAGAGGAAAGCGGAGCGC 1191
DB 1215 GAGGCTGAACGCAAGCCAGAAAGAGAAAGAGTGGAGGCGGAAACAGAGAACTG 1274
QY 1192 CAGAGACAGAGGCAAGCGGCACTGTGAGCTGAGAGACGACGTGAGAAAGCAGCGGAG 1251
DB 1275 CAGAGCAAGATGGAAGAGAGAGCTGAGAGTGTGAGAAACGCTTGGAGAAACAGAGAG 1334
QY 1252 CTGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGATGAGAGGCGGAGCGGCAAA 1311
DB 1335 CTGAGAGAGAGCGGAGAGAGAGAGAGAGAGAGAGATGAAAGATGAAAGAGAGGAGCAAA 1394
QY 1312 CCGGAACTGGAAGAGGAGGAGCAACTTGAATGGGAAACGGAACCGAGACAGGAACTCTG 1371
DB 1395 CAGAGCTTGAAGAGCAAGCGCGTTTGAATGGGAAAGACTCGTGGCAGAGAGCTGCTC 1454
QY 1372 AATCAGAGAAACAGAGAGAGAGAGGCAACCTGTGCTGAGAGCAAGAGAGAACTCTG 1431
DB 1455 AGTCAAGAAACAGAGAGAGAGAGAGATTTGACAGCTGAGCTCCAGAAAGAAAGTCTC 1514
QY 1432 GAGTTGATGAGAGCTGGAATGACAAAGACATCAGCTGAGAGAGAAACCTTCAGAGAT 1491
DB 1515 CACCTGGAATGGAAGAGAGAGAGAGAGAGAGAGAGATCTCAGAGCAGATCAAGAT 1574
QY 1492 ATCAGGTGCGACTGGCAACCCAGAGGCAAGAAATTTGAGAGACGAAACAAGTCTAGAG 1551
DB 1575 GTCAAAATGAGAAAGAAACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1634
QY 1552 CTAAAGATTTGGAATTCACCACTTACAGCAGAGTTGACAGAAATCTCAGCAATGCTT 1611
DB 1635 CTGAAATTTATGGAATTCAAACACTTCAACAGAGCTTAAAGAAATATCAAAATTAAGCTT 1694
QY 1612 GGAAGACTTATCCAGAGAAAGAGATATCTCAGTACAGTTAAACAATCCAGAGAGAC 1671
DB 1695 ATCTATCTGCTCCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1754
QY 1672 AGTTGCTATAGAGACTCGCTTCTTACCCCTCAAAAGAGCTTGAAGAGAGAGAGAGAG 1731
DB 1755 AACACAGCTGATTCAGAGGATCAGTTTATCTTCAATTAAGTCAATCAAGAAAGAGAGATTA 1814
QY 1732 CCGCAGAGAGCTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1791
DB 1815 TCCCAAGAGCTTAAAGAACATTAATGAGTCTTTGAAAAAGAAAGTCACTTAAGCTCTCA 1874
QY 1792 GAGATTTGATGTTTCAACAACAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1851
DB 1875 GAAATGAGATTCATTTAACAATGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1934
QY 1852 CTCAGAGAGAGAGTCCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1911
DB 1935 TTAGGCTTGAACAACCTCATTAATTAACAGCTGACAAATTTGAGAGAAATGAGAAAGAAA 1994
QY 1912 AGCTGAGAGTTGAGAGAGAGAA 1934
DB 1995 AGATTGAGAGCAAAAAA 2017

Db 1815 TGCCAAAGACTTAAAGAACATTAAGTCTCTTGAAGAAAAAAGCATCTAAGCTTCA 1874
Qy 1792 GAGATTGATGTTTTCACAAACAGGCTGAAGAACTGAGAGATCATAGCAACAGCA 1851
Db 1875 GAAATGATTTCAATTAACATAGCTGAAGAACTCAGAGAAAGCTATATACAGCAG 1934
Qy 1852 CTCGAAGACAGAGGCTCCCTGAGGCGAGCACTGAGACAGAAAGAGAGAGAG 1911
Db 1935 TTAGGCTTGAACACTTATTAATCAAAAGCTGACAAATGAGAGAAATCGAAGAAA 1994
Qy 1912 AGCTGAGATTGAGAGCAAAA 1934
Db 1995 AGATTAGACAAAAA 2017

RESULT 4
US-10-198-053-72
; Sequence 72, Application US/10198053
; Publication No. US20030124140A1
; GENERAL INFORMATION:
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Rafter, Gary R.
; APPLICANT: Hill, Paul
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.462C9
; CURRENT APPLICATION NUMBER: US/10/198,053
; NUMBER OF SEQ ID NOS: 624
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 72
; LENGTH: 2017
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-198-053-72

Query Match 13.1%; Score 477.4; DB 14; Length 2017;
Best Local Similarity 55.5%; Pred. No. 8.6e-127;
Matches 1111; Conservative 0; Mismatches 796; Indels 96; Gaps 6;

Qy 1 ATGGCTCAGTTTCCACACCTTTCGGTGTAGCTGATGCTGGCCATTAAGTGAG 60
Db 42 ATGGCTCAGTTTCCACACGATGATGAGAGGCCAAATATGTGGCTATTAATCTGA 101
Qy 61 GAAAGGCGCAAGATGACGAGATCTTACGCTGAGCCGATAGCGGATTTATTA 120
Db 102 GAACTACTAAGCATGATTAACAGTTTGAATACCTCAAACTTCAGAGATTACATA 161
Qy 121 GGTGATCAAGGAGGAACTTTTTCATCTGCTTACCTCAAGCTGTCTTACAGCA 180
Db 162 GGTGATCAAGGCGGATCTTTTTCATCTGCTTACCTCAAGCTGTCTTACAGCA 221
Qy 181 ATATGCGCGCTATCGGACATGATTAAGATGAGATGATCAAGTGAATTTTCA 240
Db 222 ATATGCGCGCTTATCATGATGATTAAGATGAGATGATGATCAAGTGAATTTTCA 281
Qy 241 GCGATGAAGCTTATCAAACTGAGCTACAGGATATCACTCCCTCCACACTTCCCT 300
Db 282 GCTATGAAGCTATCAAGTTAAAGTTGACGAGGCCAAACAGCTCTGATGCTCTCCCT 341
Qy 301 GTCATGAAGACGACCAAGT---GCTATTTCAAGTCAAGCATTTGATTAAGAGG 357
Db 342 ATCAATGAAGACCAAGCTTATGTTCTCTCACTAATCTGCTGTTTGGAGTGAAGC 401
Qy 358 ATTTGCTAGCATG-----CCACCACTCAGAGCTGTTGCTCTGTGCGCA----- 400
Db 402 ATGGCCAAATCTGCTCATTCATGAGCATTTGCTTCAAGTTGACATATGACCAACCTTG 461
Qy 401 -----TGGGCTTCATTCAGTTTGGATGTCTCACCTTAATCT 444
Db 462 TCTTCTGCTACTTCAAGGACAGATATCTCTCCCTAATGATGCTGCTCTTACGCT 521

Qy 445 TCTGTCCCTCAGCAGCAGCTGCTCCCTGGCTAAAGGGGCTCTCCGCTACAGCCT 504
Db 522 TCTGTATGATCACTCTCTATTAACAATGAGATGCTGCTATTCATTCAGCTTATCATT 581
Qy 505 CTGCTGCGTTTGGCATCTGACGCCACATGCGCAAGAGTCTTCTTCCAGCATCT 564
Db 582 CTTATTTCTTCTTCAACATGCTGATCATCTTATCAGCTGATGAGAGATTT 641
Qy 565 GGTCCAGGATCAATTAACATTAAGTACAGAGCAATCAATTGATGTCCCGCAG 624
Db 642 GGTGTGTAGTATTCAGAGGCCCAATCTGATGATTAAGATCTAGTATGATCTCACT 701
Qy 625 GCCCTCCAG-----CAGCAAAATGGGCTGTG 651
Db 702 TCTTAATCTGCTCTCTCAGGAACTCAGTAAAGAGAGAGAGAGAGAGAGAGAG 761
Qy 652 CTTCAATCAAGGCTGAAATACAGGAGTTTCAACAGCAGCAAGCAAACTATGAGT 711
Db 762 CTTCAAGCTTCAAGTTAAAGTATGCGCAAAATTAATGATCTAGACAAAGGATGAGC 821
Qy 712 GACATCTTAACAGGCTCCCGAGCAAGAACTATCTCATGCAATCAAGTTTACCCAGGCT 771
Db 822 GATTAATCTTCAAGTTTCAAGTTAAAGTATGAGTATGAGTATGAGTATGAGTAT 881
Qy 772 CAGCTGCTCAATATGAGATCTTCTGATGATGAGAGGCTGCTGAGAGCGGCTGAGAT 831
Db 882 CAGCTAATCTATTTGATGATCTGCTGATGATGATGATGATGATGATGATGATGAT 941
Qy 832 GAATTTATCTAGCTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 891
Db 942 GAATTTATCTGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1001
Qy 892 GTCTGCTCCAGATATATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 951
Db 1002 AGTTGCTCTCCAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1050
Qy 952 TCCGTATAGCTCTTCTCTGATGATGAGAGGCTGCTGAGAGCGGCTGAGAT 1011
Db 1051 -----TTGATTTCTGTTATGAACTGCTCTTATCAATTAAGAAACAGAA 1097
Qy 1012 GAGCAGCAGCAGAGAGAACTGCTGATGATTTGAATTAAGAACTGAGAGAACTTC 1071
Db 1098 GA---AGAGCTCAGAAAGAACTGCAATTTCTTTGAGCAAAAGCAAGCAATCTAT 1154
Qy 1072 GAGCAGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1131
Db 1155 GAACTGAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1214
Qy 1132 GAGCAGAGCGGTTGGCTCAGCTGAGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAG 1191
Db 1215 GAGGCTGAACGAAAGCCAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1274
Qy 1192 CAGGAGCAGAGGCGCAAGCTGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1251
Db 1275 CAGAGCAAGATGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1334
Qy 1252 CTGAGCGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1311
Db 1335 CTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1394
Qy 1312 CGGGAATGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1371
Db 1395 CAGAGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1454
Qy 1372 AATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1431
Db 1455 AGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1514
Qy 1432 GATTTGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1491
Db 1515 CACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1574
Qy 1492 ATCAGGTTTCACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1551

Db 1575 GTCCAAATCGAAGAAAGCAAAACACAAAAGCTGAAGCTTGGATTAACAGGTGC 1634

Qy 1552 CTAAGAATTGCTGAATCACCACCTTACAGACAGGTCAGAAATCTCAGCAATGCTT 1611

Db 1635 CTGGAATAATTAGAAATCAAAACAATTCAACAAGAGCTTAAGAAATATCAAAATAGCTT 1694

Qy 1612 GGAAGACTTAATCCAGAGAAAACAGATCTCAGTACCCAGTTAAAAACAAGTCAGCAGAAC 1671

Db 1695 ATCTATCTGGTCCCTGAGAGACAGGTATTAAACGAAAGATTTAAACATCAGCTCGT 1754

Qy 1672 AGTTGCATAGAGA CTGGCTTCTTACCCTCAAAAGAGCCTTGAAGCAAGAGCTGACC 1731

Db 1755 AACACACCTGATTACGGGATCAGTTTACTTCAATAAAAAGTCATCAGAAAAGAGAAATTA 1814

Qy 1732 CGGACAGCAGCTCCGGAGACAGCTGGACGAGGTGAGAGAGACACAGGCTCAAGCTGCAG 1791

Db 1815 TGCCAAAGACTTTAAAGAACAAATTAGTCTCTTGAAAAAGAAACCTGCATCTAAGTCTCA 1874

Qy 1792 GAGATTGATGTTTTCACCAACCAAGCTGAGGAATCTAGAGAGATACATAGCAACAGCA 1851

Db 1875 GAAATGATTCATTTTAAACAATCAGCTGAAGGAATCAGAGAAAGCTATATATACACAGAG 1934

Qy 1852 CTCACAGAGCAGAGGTCCTCTGAGGACAGCGGCACTGAAGCAAGAAAGACAGAGAGAG 1911

Db 1935 .TTAGCCCTTGAACAACCTTCATTAATCAAAAGTGAACAATTTAGAGGAATCGAAAGAAA 1994

Qy 1912 AGCCTGAGTTAGAGAACGCAAAA 1934

Db 1995 AGATTAGACAAAAAAGAAAAA 2017

RESULT 5

```

US-09-764-881-55
Sequence 55, Application US/09764881
Publication No. US20030125245a9
GENERAL INFORMATION:
APPLICANT: Rosen et al.,
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PT207
CURRENT APPLICATION NUMBER: US/09/764,881
CURRENT FILING DATE: 2001-01-17
Prior application data removed - refer to PALM or file wrapper
NUMBER OF SEQ ID NOS: 192
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 55
LENGTH: 568
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (481)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (536)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (556)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (562)
OTHER INFORMATION: n equals a,t,g, or c
US-09-764-881-55

```

Query Match	9.8%;	Score 356;	DB 11;	Length 568;
Best Local Similarity	86.2%;	Pred. No. 5.2e-92;		
Matches 426;	Conservative	0;	Mismatches 64;	Indels 4;
			Gaps	3;

Qy 1 ATGCTAGATTGTTCCACACCTTTCCGTGGTACCTGATGTCGTGGCCATAACCTGTGGAG 60
Db 78 ATGCTAGATTGTTCCACACCTTTCCGTGGTACCTGATGTCGTGGCCATAACCTGTGGAG 137
Qy 61 GAAAGGGCCAAAGCATGACGACGACCTTTAGCTGAAAGCCGATAGCGGGATTTATTACT 120

Db 138 GAAAGGCCA--CATGATCAGAGTTCATAGTTTAAAGCAATATCTGGATTCATTACT 195

Qy 121 GGTGATCAAGCAGAGAACTTTTTTCCAAATCTGGGTTACTCAAGCTGTCTTAGCACAA 180

Db 196 GGTGATCAAGCTAGAAACTTTTTTTCAAATCTGGGTTACTCAAGCTGTCTTAGCACAG 255

Qy 181 ATATGGGCCCTACCGGACATGAAATAACATGGAAGGATNGATCTCAAGTGAATTTTCATA 240

Db 256 ATATGGGCACTAGCTGCACATGAATATATGGAAGAAATGATCAAGTGGAGTTTCCATA 315

Qy 241 GCCATGAGCTTATCAAATCTGAAGCTACAGAAATATCAGCTCCCTCACACTTCCCT 300

Db 316 GCTATGAATCTTATCAAATCTGAAGCTACAGAAATATCAGCTCCCTCACACTTCCCT 375

Qy 301 GTCATGAAACAGCAACCACTGGCTATTTCAATGCAACCAAGCATTTGGTATAGAGGATT 360

Db 376 GTCATGAAACAGCAACCACTGGCTATTTCTAGCGACAGCATTTGGTATGGAGGATTC 435

Qy 361 GCTAGCATGCACCACTCAAGCTGTTCCTCTGTGCCAATGGGCTCAATCCAGTTGT 420

Db 436 GCCAGCAAGCACACCGCTTACAGCTTTCCTCAGTGCCAATGGGAGNCCATTCAGTTGT 495

Qy 421 -GGAAATGTTCCACCTTAGTATCTTGTGTCCCTCCAGACAAATGGCTCCCTGGGCTAA 479

Db 496 GGGAAATGTTCCCAACCTTAGTATCTTGTGTCCACAGACA-NTGTGCCCCCTCGGCTAA 554

Qy 480 CGGAGGCTCTCCCG 493

Db 555 AAGGGATNCCCTTG 568

RESULT 6

US-09/879-957-193
Sequence 193, Application US/09879957
Patent No. US20020034755A1
GENERAL INFORMATION:
APPLICANT: SPARKS, Andrew B.
HOFFMAN, No. US20020034755A1
KAY, Brian K.
FOWLES, Dana M.
McCONNELL, Stephen J.
TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
USING SAME
NUMBER OF SEQUENCES: 227
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/879,957
FILING DATE: 13-Jun-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/630,915
FILING DATE: 03-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-174
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 193:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2873 bases
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: unknown
 MOLECULE TYPE: DNA
 SEQUENCE DESCRIPTION: SEQ ID NO: 193:
 US-09-879-957-193

Query Match 9.1%; Score 332.4; DB 9; Length 2873;
 Best Local Similarity 61.0%; Pred. No. 9e-85;
 Matches 631; Conservative 0; Mismatches 361; Indels 42; Gaps 4;

```

QY 2609 CTTCTCTGACCGTACTAGTGTGGCCAGTTACGGAGAGATCAGCTTTAATCCCGGCCA 2668
DB 541 CTTTTCCTGACCGTACTAGTGTGGCCAGTTACGGAGAGATCAGCTTTAATCCCGGCCA 600
QY 2669 CAGCCACTGGCTCTCCCTCCATCTCCGCTCTGGCCAGGGTGAAGAGTGAAGGGCTAC 2728
DB 601 TGTCCCTGG---ATCTGTATACCTATTCATGAGACAGGAGACAGTGTGAGAACTTAA 657
QY 2729 AAGCCGACCCCTGTATCTCCCTGAGAGCCAAAGAGACACCACTTAAATTTTACAAA 2788
DB 658 AAGCAGAGCCCTTGTCTCTGAGCTGCAAGAAAGATTAACCACTTGAATCTCTCAAAAC 717
QY 2789 GTGAGCTATCACCGTCTGGAACAGACAGACATGTGTGTGGTGTGGAGAACTTCAAGTC 2848
DB 718 ATGACATTATATGCTTGTGAGACAGCAAAATGTGTGTGGAGGTGATGAGAG 777
QY 2849 AGAAGGTTGTGTCTCCCAAGTCTTACGTAAATCTCATTCAGGGCCCTTGAAGAAATCA 2908
DB 778 GAAAGAGATGTTTCCCAATCTTATGTCAAGTCAATCTCTGGAGATGAGTAAACGGG 837
QY 2909 CAGACATCATCTGCGCCCTTCTGAAAGTCTCTGATGCTTAAAGAGAGTGGCTTCCCGG 2968
DB 838 AA-----GAAACAGAAAGCTTGTGTATGACCTGTAAATAGAAACCTACCTCGG 885
QY 2969 CCGCCAGGACGACATCCCGGAGAGAGATTTATGTCATATGACATATCGAGATCTCG 3028
DB 886 CAGCTTATTCAGT-----GGAGAAAGATATATATGACCTTATTCATATTCAGAGTGG 939
QY 3029 AGCAAGAGATTTAACTTTTACAGCAAGGGAGTGTGTGTGTGTTACCAAGAAAGATGAG 3088
DB 940 AACCTGAGATTTGACTTTTACAGAGAGTGTGAAATATTTGTGACCCGAAAGATGAG 999
QY 3089 ACTGTGAGCGGAAAGCGTGGCGACAGTCCGAGCTTCTCTTAACTATGTAGGC 3148
DB 1000 AGTGTGAGCAGAGAGATTTGAGAGATGAGATGAGATTTTCCATCAATCTATGTCAAC 1059
QY 3149 TTAAGATTCAGAGGCTCTGGAACCTGTGGGAGAAACAGGAGTTTAAAGAAAAAATCTG 3208
DB 1060 CAAAGGATCAAGAGATTTTGGAGGTGAGCAAGTGTGAGCCTCAATTAATAAATCTG 1119
QY 3209 AATTGCGCAGTATTTGCTTCTTACGCTACTGCTACCTCCGAAACACTCACCTGGCTC 3268
DB 1120 AATTGCTAGGTTACTTACAGATATGTTGCTTCTGCTTCTGAAACACTTAAAGCTTGGC 1179
QY 3269 CTGGGAGCTGATCTGATCTCGGAAAGAAACCCAGTGTGATGTGTGAGAGAGAACTGC 3328
DB 1180 CAGGACAGTATTAATTAATTTAAAGAAATATCAAGTGTGTGTGAGAGAGAGTTCAC 1239
QY 3329 AAGCTGAGGAGAAAAAGCCGAGATGAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3388
DB 1240 AAGCAGAGAGAAAAAGCAGAGAAAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1299
QY 3389 GCCCGGAAACAGCAAAATCACCCCACTGAGCTACCAAGACCGAGTGCAGCCAGCAG 3448
DB 1300 GTCCAGATGTAAGAGAGCAGCCTGCTTTTATC-----CTG 1338
QY 3449 TGTGCGAGGTGATCGGATGTATCGATTACCGCCGAGAGAGATGAGAACTGACCTTCA 3508
DB 1339 TATGTGAGGTGATGCTATGTATGACTATGACAGCAAAATATGAGATGAGTCAAGTTC 1398
  
```

```

QY 3509 GCAAGGCCAGATCATCAAGTCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3568
DB 1399 CAAAGGACACTCTTATTTATGTAAGAAAGATGATCTATGTTGTGTGCAAGAGAGA 1458
QY 3569 TCAGTGGGCAAGTGGGCTCTTCCATCCATTAATGATTAAGAGTGCACAGACATGACC 3628
DB 1459 TCAAGGGGTGACTGTGTCTTCTTCAACTAGATTAAGATGAGACAGACTCAGATC 1518
QY 3629 CAGCCAGCAATGA 3642
DB 1519 CAAGTCAACAGTGA 1532
  
```

RESULT 7

```

US-09-764-868-125
; Sequence 125, Application US/09764868
; Patent No. US2002016871A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT232
; CURRENT APPLICATION NUMBER: US/09/764,868
; PRIOR APPLICATION DATA: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 125
; LENGTH: 4210
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-868-125
  
```

Query Match 9.1%; Score 331.4; DB 10; Length 4210;
 Best Local Similarity 61.0%; Pred. No. 2.2e-84;
 Matches 630; Conservative 0; Mismatches 361; Indels 42; Gaps 4;

```

QY 2609 CTTCTCTGACCGTACTAGTGTGGCCAGTTACGGAGAGATCAGCTTTAATCCCGGCCA 2668
DB 918 CTTTTCCTGACCGTACTAGTGTGGCCAGTTACGGAGAGATCAGCTTTAATCCCGGCCA 977
QY 2669 CAGCCACTGGCTCTCCCTCCATCTCCGCTCTGGCCAGGGTGAAGAGTGAAGGGCTAC 2728
DB 978 TGTCCCTGG---ATCTGTATACCTATTCATGAGACAGGAGACAGTGTGAGAACTTAA 1034
QY 2729 AAGCCGACCCCTGTATCTCCCTGAGAGCCAAAGAGACACCACTTAAATTTTAAACAAA 2788
DB 1035 AAGCAGAGCCCTTGTCTCTGAGCTGCAAGAGAAAGATTAACCACTTGAATCTTCAAAAC 1094
QY 2789 GTGAGCTATCACCGTCTGGAACAGCAAGATGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2848
DB 1095 ATGACATTATTAATCTGTGTGAGCAGCAAGAAATATGTGTGTGTGTGTGTGTGTGTGT 1154
QY 2849 AGAAGGTTGTGTCTCCCAAGTCTTACGTGAATCAATTTCAAGGCGCCGTAAGAAATCA 2908
DB 1155 GAAAGAGATGTTTCCCAATCTTATGTCAAGATCATCTCTGGAGTGAAGTAAACGGG 1214
QY 2909 CAAAGTCAATCTGAGCTTCTGAAAGTCTGTAGTCTTAAAGAGTGTGCTTCCCGG 2968
DB 1215 AA-----GAAACAGAAAGCTTGTATGAGCAGTGTAAATTAAGAAACCTACCTCGG 1262
QY 2969 CCGCCAGGACGACATCCCGGAGAGAGATTTATGTCATGATGACATATTCAGAGATCTG 3028
DB 1263 CAGCTTATTCAGT-----GGAGAAAGATATATTTGCACTTATTCATATTCAGATGTGG 1316
QY 3029 AAGCAGAGATTTAACTTTTACAGCAAGGGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3088
DB 1317 AACCTGAGATTTGACTTTTACAGAGAGTGTGAAATATTTGTGTGACCCAGAAAGATGAG 1376
QY 3089 ACTGTGAGCGGAAAGCGTGGCGACAGTCCGAGTCTTCCCTTCAACTATGTAGGC 3148
DB 1377 AGTGTGAGCAGAGAGATTTGAGAGATGAGATGAGATTTTTCATCAAACTATGTCAAAC 1436
  
```

OY	3149	TTAAAGATCTAGAGGGCTCTGAACTGCTGGGAAAAACAGGAGTTTACGAAAAAAACCTG	3208
Db	1437	CAAGGATCAAGAGGTTTGGGAGTCTACGAACTCTGAGCATTCAATATAAAAACCTG	1499
OY	3209	AAATTGCCAGGTTATTGCTTCTTAGCTGCTACTGTCCGAAACACTACCTTGCTC	3268
Db	1497	AGATTGCTCAGGTAACTTCAGCATATGTTGCTTCTGGTTCTGAACTTAGCCTTGGAC	1556
OY	3269	CTGGGACAGCTGATTCGTATCCGGAAAAAAGAACCCAGGTGAGATGGTGGAGAGAGATCG	3328
Db	1557	CAGACAGTTATATTAATTCTAAAGAAAAATACAGATGGGTGGTGGCCAGAGAGATTAC	1616
OY	3329	AAGCTCGAGGAAAAAGCCGACAGATAGGTGGTTCCAGCAATTATGTCAACTTCTAA	3388
Db	1617	AGGCGAGAGGAAAAAGGACAGAAAGGATGGTTTCTGCCAGTCATGTTAACTTTTGG	1676
OY	3389	GCCCCGAAACAAGCAAAATCACCCCCACTGAGCTACCCAGACCGCAGTCGACGACAG	3448
Db	1677	GTTCCAAATGATGAAAGAGCCACACTCTGCTTTATC-----CTG	1715
OY	3449	TGTGTCAGGTATCGGAGTGTACGATTACACCGCCAGACGATGACGAACTTTC	3508
Db	1716	TATGTCAGGTATGCTATGTATGACTACTGACGAAATATAGAGATGAGCTCAGTTCT	1775
OY	3509	GCAAAGCGACATCACTCAACGCTCCCAACAGAGAGACCCGAGCTGTGSAAGAGAGAG	3566
Db	1776	CCAAAGGCACTCATATATGTTATGAACAAAGATGATCTGATTTGGGCAAGAGAGA	1835
OY	3569	TCAGTGGGCAAGTTGGGCTCTTCCCATCCAAATTATGTAAGCTGACCAAGACATGAGC	3628
Db	1836	TCAACGGGGTAGCTGCTCTCTTCTTCCAACTACGTTAAGATGACGACAGCTCAGATC	1895
OY	3629	CCAGCCAGCAATG	3641
Db	1896	CAAGTCAACAGTG	1908

RESULT 8

```

Sequence 31258, Application US/09918995
Publication No. US20030073622A1
GENERAL INFORMATION:
APPLICANT: Hyseq, Inc.
TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
FROM VARIOUS CDNA LIBRARIES
FILE REFERENCE: 20411-756
CURRENT APPLICATION NUMBER: US/09/918,995
PRIOR FILING DATE: 2001-07-30
PRIOR APPLICATION NUMBER: US/09/235,076
PRIOR FILING DATE: 1999-01-20
NUMBER OF SEQ ID NOS: 38054
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 31258
LENGTH: 503
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)...(503)
OTHER INFORMATION: n = A,T,C or G
US-09-918-995-31258

```

Query Match	8.9%	Score 322.6;	DB 11;	Length 503;
Best Local Similarity	81.6%	Pred. No. 2.2e-82;		
Matches 386;	Conservative 0;	Mismatches 84;	Indels 3;	Gaps 1

QY	823	ACTGAGAAAGAAATTATCTGCTAGCTAGACCCGTAATGTAATGTCGATGTCGTGAGCA	882
Db	30	AATTAGAGGAATTTATCTGCGCAATGCACTCATTTGATGTGATGCTATGCTGCGCAACCA	89
QY	883	CTGCCGCCGCTCTGCTCTCCAGAAATACATCCCTCTCTCTTCAAGAAATTGCTCCGGC	942
Db	90	CTGCCACCTGTCTCTGCTCCAGAAATACATTCACCTCTTTTAAAGAAATTGATCTGGC	149

Qy	943	AGTGGATATGCCGATTAAGCTCTTCTCTGTGGATCAGAGGCTGCTGAGAGACCGTG	1002
Db	150	AGTGTATATCTGTCAATTAAGCTCAACATCTGTGATTAAGGCTTACCAAGAGAAACCA	209
Qy	1003	TCAGAGATGAGCAGCAGC---CAGAGAGAAATCGCTGTGACATTTGAAGATTAAG	1059
Db	210	TTAGAAATATGACACAAACAATTAGAAABAATAATCTGTAACTTTGAAGATTAAG	269
Qy	1060	CGGAGAACTTTGAGCGAGGCGAGTGTGAGCTGAGAAACGCGCCAAAGCCTTTGAG	1119
Db	270	CGGAGAACTTTGAACTGTGCAACTGTGAACTGTGAGAAACGAAAGCAAGCTCTCTG	329
Qy	1120	CAGTAGGCGAAAGAGCAGAGCCGTTGGCTCAGCTGAGAGCGCGCCGACAGAGAA	1179
Db	330	CAGTAGGCGAAAGAGCAGAGCCGCTTGCCACGCTGAGCGGCGGAGCAGAGAAAG	389
Qy	1180	GAGCGGAGGCGCCAGAGAGCAGAGGCGCAAGCGGAGCTGAGCTGAGAAAGCAGTGG	1239
Db	390	GAGGTTAGCGCCAGAGAGCAAGACCGCATTAAGACACTTGAATCTGGAGATGCACTGG	449
Qy	1240	AAGCAGCGGAGCTGAGCGGCGAGCGAGAGAGAGAGAGAAAGAGATCGA	1292
Db	450	AAGCAGCGGAGCTGAGAGCGCAGAGAGAGAGAGAGAGAAATATTGA	502

RESULT 5

US-09-879-957-39
Sequence 39, Application US/09879957
Patent No. US20020034755A1
GENERAL INFORMATION:
APPLICANT: SPARKS, Andrew B.
HOFFMAN, No. US20020034755A1h
KAY, Brian K.
FOWKES, Dana M.
McCONNELL, Stephen J.
TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
USING SAME
NUMBER OF SEQUENCES: 227
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/879,957
FILING DATE: 13-Jun-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/630,915
FILING DATE: 03-Apr-1996
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-174
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 39:
SEQUENCE CHARACTERISTICS:
LENGTH: 747 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

OY 2539 AGTCCACGT 2548
|||
Db 241 AGTCCACGT 250

RESULT 11
US-09-864-761-26948
Sequence 26948, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecolica-X-1
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US/09/864,761
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 26948
LENGTH: 286
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AP000311.1
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 0.87
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.3
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 32
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.66
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.69
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.92
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.98
OTHER INFORMATION: NT HIT: AF114488.1, EVALUATE 0.00e+00

OTHER INFORMATION: SWISSPROT HIT: Q15811, EVALUATE 2.00e-45
OTHER INFORMATION: EST_HUMAN HIT: AA773263.1, EVALUATE 1.00e-112
US-09-864-761-26948
Query Match 5.5%; Score 198.8; DB 9; Length 286;
Best Local Similarity 87.2%; Pred. No. 1e-46;
Matches 218; Conservative 0; Mismatches 32; Indels 0; Gaps 0;
OY 2299 GTGATGAAAGCCGACTGAGAGCCAGATGCTTGGAGAGAGCTGAAAGGAAAGC 2358
|||
Db 1 GTGATGAAAGCCGAACTGGAGAACCCGCTGCTTGGAGAGATTAAAGGAAACA 60
OY 2359 GGATGTTCCCTGCAAACTATGACAAAGATTCCAGAAATGATGCTCCACTCCAGCC 2418
|||
Db 61 GGGTGTCTCCCTGCAAACTATGACAAAGATTCCAGAAATGATGCTCCACTCCAGTG 120
OY 2419 AACCAGTGAACCGATGCACTGCCCCCTGCCCCCAACTGCTGCTGAGACCCCT 2478
|||
Db 121 AACCACTGACTGATTCACATCTGCCCCCTGCCCCCAACTGCTGCTGAGACCCCT 180
OY 2479 GCTCCTTTGCAAGTACTCTTTCAGAGCCCTCCCAACCCCAACACTGCGGAGACTTC 2538
|||
Db 181 GCCCTTTGGCAGTAACTCTTTCAGAGCCCTCCCAACCCCAACACTGCGGAGACTTC 240
OY 2539 AGTCCACGT 2548
|||
Db 241 AGTCCACGT 250
RESULT 12
US-09-864-761-30453
Sequence 30453, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecolica-X-1
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US/09/864,761
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30

```

PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 30453
LENGTH: 297
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AP000117.1
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.7
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.4
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.2
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 0.76
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.3
OTHER INFORMATION: EXPRESSED IN PETAL LIVER, SIGNAL = 1.5
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.95
OTHER INFORMATION: NT HIT: AF114487.1, EVALUE 0.00e+00
OTHER INFORMATION: EST_HUMAN HIT: A473263.1, EVALUE 1.00e-112
OTHER INFORMATION: SWISSPROT HIT: Q15811, EVALUE 2.00e-45
US-09-864-761-30453

```

```

Query Match
Best Local Similarity 5.5%; Score 198.8; DB 9; Length 297;
Matches 218; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 2299 GTGATGAAGCAGACGAGAGGAGGATGCTTGAAGAGAGCTGAAGAGAGAGG 2358
DB 1 GTGATGAAGCAGACGAGAGGAGGATGCTTGAAGAGAGCTGAAGAGAGAGG 60
QY 2359 GGATGCTTCCGCAAACTATGAGAGAGGATGCTTGAAGAGAGG 2418
DB 61 GGATGCTTCCGCAAACTATGAGAGAGGATGCTTGAAGAGAGG 120
QY 2419 AAACGAGTACCGATGATGATGCTTCCGCAAACTATGAGAGAGG 2478
DB 121 AAACGAGTACCGATGATGATGCTTCCGCAAACTATGAGAGAGG 180
QY 2479 GCTCCTTTCAGTACGCTTCTGAGCCCTCCAGAACCCCACTGAGAGAGG 2538
DB 181 GCTCCTTTCAGTACGCTTCTGAGCCCTCCAGAACCCCACTGAGAGAGG 240
QY 2539 AGTTCACGT 2548
DB 241 AGTTCACGT 250

```

```

RESULT 13
US-09-864-761-17146
Sequence 17146, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecmics-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
PRIOR FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6

```

```

PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 17146
LENGTH: 301
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AP000311.1
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.2
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 3.2
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.6
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.3
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.2
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.4
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.7
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.2
OTHER INFORMATION: NT HIT: AF114487.1, EVALUE 3.50e-01
OTHER INFORMATION: SWISSPROT HIT: Q35601, EVALUE 1.00e-122
OTHER INFORMATION: EST_HUMAN HIT: BE542917.1, EVALUE 4.00e-62
US-09-864-761-17146

```

```

Query Match
Best Local Similarity 5.5%; Score 198.6; DB 9; Length 301;
Matches 228; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 1040 TGACATTGTAATAGAGGAGGAGAACTTGAAGAGGAGTGTGAGAGG 1099
DB 1 TAACTTTGAAGATAGAGGAGGAGAACTTGAAGAGGAGTGTGAGAGG 60
QY 1100 GCCGCAAGCGCTCTTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1159
DB 61 GAGGCAAGCTCTCTTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 120
QY 1160 GCGCCAGAGAGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1219
DB 121 GGGCGAGAGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 180
QY 1220 AGCTGAGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1279
DB 181 AACTGAGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 240
QY 1280 GAAAGAGATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1316

```

Db 241 GGAAGAAATTGAGAGCGGAGGTAACGACGCGGAGA 277

RESULT 14
US-09-864-761-17644
Sequence 17644, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aeomica-X-1
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 17644
LENGTH: 263
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO APO00049.1
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.3
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.2
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.3
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 3
OTHER INFORMATION: EXPRESSED IN HEL100, SIGNAL = 2.9
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.2
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.5
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 5.3
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.9
OTHER INFORMATION: EST_HUMAN HIT: BE542917.1, EVALUATE 3.00e-62

OTHER INFORMATION: SWISSPROT HIT: O35601, EVALUATE 2.80e-01
OTHER INFORMATION: NT HIT: AF114487.1, EVALUATE 1.00e-122
US-09-864-761-17644

Query Match 5.4%; Score 197.4; DB 9; Length 263;
Best Local Similarity 84.4%; Pred. No. 2.5e-46;
Matches 222; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 1040 TGACATTGGAAGATTAAGAACGGGAGAACTTCAGCGAGGACGAGTGTGAGCTGGAGAGC 1079
Db 1 TAACTTTGAAGATTAAGAACGGGAGAACTTTGAACCTGGCAACTGGAGGAAC 60
QY 1100 GCCCGCAAGCGCTCTTTGAGCAGCAGCGCAAGAGCAGAGCGGTGCTCAGCTGAGC 1159
Db 61 GAAGCAAGCTCTCTCGAAGCAGCAGCGCAAGAGCAGAGCGCGCTGGCCAGCTGAGC 120
QY 1160 GCGCGGAGCAGAGAGAGAAAGCGGAGCGCCAGAGCAGAGAGGCCAGCGGAGCTGG 1219
Db 121 GGGCGGAGCAGAGAGAGAGAGAGAGAGCTGAGCGCCAGAGCAAGAGCCGAAAGACACTGG 180
QY 1220 AGCTGAGAGAGCAGCTGAGAGAGAGAGAGCTGAGCGGCGGAGAGAGAGAGAGAG 1279
Db 181 AACTGAGAGAGCAACTGAGAAAGCAGCGGAGAGCTAGAACGAGAGAGAGAGAGAG 240
QY 1280 GGAAGAGATCGAGAGCGCGAG 1302
Db 241 GGAAGAAATTGAGAGCGCGAG 263

RESULT 15
US-09-864-761-10314
Sequence 10314, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aeomica-X-1
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US/09/864,761
PRIOR FILING DATE: 2001-02-04
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30

```
/ PRIOR APPLICATION NUMBER: US 60/234,687
/ PRIOR FILING DATE: 2000-09-21
/ PRIOR APPLICATION NUMBER: US 09/608,408
/ PRIOR FILING DATE: 2000-06-30
/ PRIOR APPLICATION NUMBER: US 09/774,203
/ PRIOR FILING DATE: 2001-01-29
/ NUMBER OF SEQ ID NOS: 4917
/ SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
/ SEQ ID NO 10314
/ LENGTH: 480
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ OTHER INFORMATION: MAP TO AP000311.1
/ OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 0.87
/ OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.1
/ OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.3
/ OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 32
/ OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.66
/ OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.69
/ OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1
/ OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.92
/ OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.98
/ US-09-864-761-10314

Query Match      5.3%; Score 192; DB 9; Length 480;
Best Local Similarity 81.6%; Pred. No. 1.3e-44;
Matches 222; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 2265 TGAGATCACCATCCAGCCAGAGATATAGTCATGTGATGAAGCCAGACTGAGAGCC 2324
DB 209 TTAGTTCTACTTAAGCTGTGACTTTTCCACAGGTGATGAAGCCAACTGAGAGACC 268
QY 2325 AGGATGGCTTGAGGAGAGAGTGAAGGAAAGAGCGATGTTCCCTGCAAACTATGACAG 2384
DB 269 CGGCTGCTTGAGGAGAAATTAAGAAAGACAGAGGTGTTCCCTGCAAACTATGACAG 328
QY 2385 AAAGATTCAGAAATGAGGTTCCCACTCCAGCCAAACAGTACCGATCTGACATCTGC 2444
DB 329 GAAATATCCAGAAATGAGGTTCCCGCTCCAGTAAACAGTATGATTCACATCTGC 388
QY 2445 CCCTGCCCCCAAACTGGCTCTGCGTGAGACCCCTGCTCTTGGCAGTGAAGCTCTTCTGA 2504
DB 389 CCTGCCCCCAAACTGGCTCTGCGTGAGACCCCGCCCTTGGCAGTGAAGCTCTTCTGA 448
QY 2505 GCCCTCCAGAACCCCAACAACTGGGCAAGACT 2536
DB 449 GCCCTCCAGAACCCCAACAACTGGGCGCAACT 480
```

Search completed: December 5, 2003, 06:15:46
Job time : 777.554 sec8

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 4, 2003, 15:21:40 ; Search time 12013.2 Seconds

(Without alignments)
17313.032 Million cell updates/sec

Title: US-09-674-237A-1

Perfect score: 5084

Sequence: 1 cggcagcagagagagtgagtgag.....gagaatcgatcaacgctt 5084

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenBank: 1: gb_ha: 2: gb_hcg: 3: gb_in: 4: gb_cm: 5: gb_ov: 6: gb_pat: 7: gb_ph: 8: gb_pl: 9: gb_pr: 10: gb_ro: 11: gb_sts: 12: gb_sy: 13: gb_un: 14: gb_vi: 15: gb_wa: 16: em_fun: 17: em_hum: 18: em_in: 19: em_mu: 20: em_om: 21: em_or: 22: em_ov: 23: em_pat: 24: em_ph: 25: em_pl: 26: em_ro: 27: em_sts: 28: em_un: 29: em_vi: 30: em_hcg_hum: 31: em_hcg_in: 32: em_hcg_oth: 33: em_hcg_mus: 34: em_hcg_pln: 35: em_hcg_rtd: 36: em_hcg_mam: 37: em_hcg_vit: 38: em_hcg_hum: 39: em_hcg_hum: 40: em_hcg_mus: 41: em_hcg_oth:

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3721.4	73.2	3723	10	AF132478 Mus muscu
2	3639.4	71.6	5145	10	AF132481 Mus muscu
3	3529.4	69.4	5381	9	AF114488 Homo sapi
4	3517.2	69.2	4025	10	AF127798 Rattus no
5	3440.2	67.7	5287	9	AF064243 Homo sapi
6	3103.6	61.0	3812	10	AF132672 Rattus no
7	2975	58.5	6439	9	AF114487 Homo sapi
8	2898.4	57.0	7247	9	AF064244 Homo sapi
9	2047.2	40.3	4103	5	AF032118 Xenopus l
10	1610	31.7	2131	6	BD127168 Primer fo
11	1610	31.7	2131	9	AK074554 Homo sapi
12	1404.2	27.6	2131	6	BD158570 Primer fo
13	1404.2	27.6	2131	9	AK027846 Homo sapi
14	1252.6	24.6	3241	9	HS061166 Human SH3 d
15	1249.2	24.6	1676	6	BD127640 Primer fo
16	1249.2	24.6	1676	9	AK075290 Homo sapi
17	1110	21.8	1133	10	AF169621 Mus muscu
18	1039.4	20.4	212305	2	AC134837 Mus muscu
19	1039.4	20.4	230097	2	AC126053 Mus muscu
20	981.8	19.3	1299	9	BC039036 Homo sapi
21	883	17.4	1996	9	AF180522 Homo sapi
22	813.6	16.0	78190	9	AC106760 Homo sapi
23	731.4	14.4	5938	2	AC023112 Homo sapi
24	731.4	14.4	5938	6	AB033082 Homo sapi
25	726.4	14.3	5828	6	BD167848 Method fo
26	726.4	14.3	5828	9	AF248540 Homo sapi
27	724.2	14.2	4557	9	AF182199 Homo sapi
28	701.2	13.8	278501	2	AC098552 Rattus no
29	701.2	13.8	278501	2	AC123507 Rattus no
30	701.2	13.8	278501	2	AC123507 Rattus no
31	679.6	13.4	135924	2	AC115529 Rattus no
32	679.6	13.3	4977	10	AF132479 Mus muscu
33	673.8	13.3	3594	10	AF132480 Mus muscu
34	659	13.0	4447	6	AX428899 Sequence
35	633.4	12.5	181782	2	AC116970 Rattus no
36	617.4	12.1	5804	10	AK122480 Mus muscu
37	611.6	12.0	877	6	BD124639 Primer fo
38	611.6	12.0	877	6	BD126208 Primer fo
39	556.2	10.9	12015	9	AP000312 Homo sapi
40	556.2	10.9	100000	9	AP000193 Homo sapi
41	556.2	10.9	114929	9	AP000050 Homo sapi
42	556.2	10.9	151516	9	AP000117 Homo sapi
43	556.2	10.9	340000	9	AP001718 Homo sapi
44	553.8	10.9	2020	9	AF064245 Homo sapi
45	540.8	10.6	184526	2	AC144367 Papio anu

ALIGNMENTS

RESULT 1
LOCUS AF132478 3723 bp mRNA linear ROD 09-MAR-1999
DEFINITION Mus musculus Esei protein mRNA, complete cds.
ACCESSION AF132478
VERSION AF132478.1 GI:4378884
KEYWORDS
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 3723)
Sengar, A.S., Wang, W., Bishay, J., Cohen, S., and Egan, S.B.
The EH and SH3 domain Esei proteins regulate endocytosis by linking
to dynamin and Eps15

Pred. No. is the number of results predicted by chance to have a

JOURNAL MEDLINE
 PUBMED 10064583
 REFERENCE 2 (bases 1 to 3723)
 AUTHORS Sengar,A.S., Wang,W., Cohen,S., Bishay,J. and Egan,S.E.
 TITLE Direct Submission
 JOURNAL Submitted (02-MAR-1999) Programs in Cancer & Blood
 Research/developmental Biology, The Hospital for Sick Children, 555
 University Avenue, Toronto, ON M5G-1X8, Canada

FEATURES
 source
 1..3723
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 1..3642
 /codon_start=1
 /product="Basil protein"
 /protein_id="AADI9746.1"
 /db_xref="GI:4378885"
 /translation="MAOPPTPGSLDVAITVERBAKHDOPLSLKPIAGPITDPA
 RNFPSQSLPOPVLAQIWMALADNNNDGMDVBSIAKLKIQYOLSTTPYM
 KOQVVISAPAFGIIGIASMPPLTAVAPVWMSIPVGMSPVLSVPVAPVPLAN
 GAPVIOPLPAFAHPAATWPKSSFSRSGPSQLNTKLQKQSFVVASAPAAEAVP
 OSSRLKYROLFNSHDKTWSGHLTGPOARTILMOSLPQALASINLSDIDODGLTA
 BEFLIAMLIDIVAMSGOPLPVLPEYIPSPRRVRSQSGMSVSSSVQDLPEPS
 SEDBOQPKKLVPTEFDKKRENERGVSLEKRRQALIEQKREBERLQLEBRAEOR
 KERERQBOBAKQLELEKQLEKREBERKEIERBAARELEROLEMER
 NRRELQNRNKEQEGIVLKAARKTLEFELALNDKHQLEKQDRLCLATQRE
 IESTKSELRILAEITLQLOQLESQMLGRLIPKQILSDQLKQVNSLHRSLLT
 LKRALEKELRLAEITLQLOQLESQMLGRLIPKQILSDQLKQVNSLHRSLLT
 EEARLKOKEBERKSLELEKQLEDAORVQERDQWLHEVQOEGEORPKPHEDELRV
 EDSVRKEAERAKPEMDKOSRLPHQEPKATLQAPWSTTEKGLTIAQESVTV
 VYRRLYPRESHDEITIQPDLYMDESQCEPQMLGELKKTGTFPANYAKIP
 ENEVTPAKPVDTLSAPAPKALRETPALPVSSEPTIPNNMADSSPSSSSNE
 KRETNDMTWAQPSLTVPASQQLQKRAFPATATGSPSPVLQGEKRVGLQAL
 YPMRAKNDHLNLFNSKSDVITVLEQDDMMFGEVQCKMFPFSYKLTISGVRKSTSI
 DTGPSPASLKRVASPAKPAIPGEEFTAMTYESSQGLDFOQGVIVTKKDG
 WMTGVGKSGVFPNSYVRLNDESGTAKGKGLPKPAQVIAQVTAAPGDELTV
 APGQILIRKRNPGMWEGLQARSKQIQIGFPANYIKLSRQSKTIPHELPRTIV
 QPAVCQVIGMTDYAQNDELAFSGQILINLKEDPDMKGEVSGVGLFSPNVKL
 TTDMP500"

CDS

BASE COUNT 1065 a 940 c 1026 g 692 t
 ORIGIN

Query Match 73.2%; Score 3721.4; DB 10; Length 3723;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 3722; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

259 ATGGCTCAGTTTCCCAACCTTGGTGTAGCTGTGAGCCATACTGTGAG 318
 1 ATGGCTCAGTTTCCCAACCTTGGTGTAGCTGTGAGCCATACTGTGAG 60
 319 GAAAGGGCCAGCATGACGAGCAGTCTTACCGTGAAGCCGATAGCGGATTTA 378
 61 GAAAGGGCCAGCATGACGAGCAGTCTTACCGTGAAGCCGATTTA 120
 379 GGTGATCAAGCGAGAACTTTTTCATCTGGGTTACCTCAGCTTCTTACGAA 438
 121 GGTGATCAAGCGAGAACTTTTTCATCTGGGTTACCTCAGCTTCTTACGAA 180
 439 ATATGGGCGCTAGCGGACATGATAACATGAGAGATGATCAAGTGAATTTTTCATA 498
 181 ATATGGGCGCTAGCGGACATGATAACATGAGAGATGATCAAGTGAATTTTTCATA 240
 499 GCCATGAACCTTATCAACTGAAGCTACAGAGATACAGCTCCCTCCACACTTCCCT 558
 241 GCCATGAACCTTATCAACTGAAGCTACAGAGATACAGCTCCCTCCACACTTCCCT 300
 559 GTTCATGAACAGCAACCACTGCTATTTTCAGTGCACACGACTTTGGTATAGAGGAT 618
 301 GTTCATGAACAGCAACCACTGCTATTTTCAGTGCACACGACTTTGGTATAGAGGAT 360
 619 GTTAGAGTCCACCACTCAGCTGTGTCTCGTGCCTAATGGGCTCCATTCCAGTTGT 678

361 GCTAGATGCGACACACCTCAGAGCTGTGCTCTGTGCAATGGGCTCATTTCAAGTTGT 420
 679 GGAATGCTTCCACACCTTGAATATCTTCTGTCTCTCCACAGCAGTGTCTCCCTGGCTAAC 738
 421 GGAATGCTTCCACACCTTGAATATCTTCTGTCTCTCCACAGCAGTGTCTCCCTGGCTAAC 480
 739 GGGGCTCTCCCTGATACAGCCCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT 798
 481 GGGGCTCTCCCTGATACAGCCCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT 540
 799 AAGAGTCTTCTTCCACAGCAGTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT 858
 541 AAGAGTCTTCTTCCACAGCAGTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT 600
 859 GCACATATCTTCAATGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT 918
 601 GCACATATCTTCAATGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT 660
 919 TCAAGGCTGAAATCAGGCAAGTTATTCACAGCAGCAAAACTATGAGTGAACCTTA 978
 661 TCAAGGCTGAAATCAGGCAAGTTATTCACAGCAGCAAAACTATGAGTGAACCTTA 720
 979 AAGAGTCTTCCACAGCAGTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT 1038
 721 AAGAGTCTTCCACAGCAGTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT 780
 1039 TCAATATGAATCTTCTGACATTCATGATGAGTGAAGAACTCAGTCGAGAAATTTATC 1098
 781 TCAATATGAATCTTCTGACATTCATGATGAGTGAAGAACTCAGTCGAGAAATTTATC 840
 1099 CTACTATGACCTAATGATGTTGTCATGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT 1158
 841 CTACTATGACCTAATGATGTTGTCATGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT 900
 1159 CCAGATATACATCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1218
 901 CCAGATATACATCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 960
 1219 AGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1278
 961 AGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1020
 1279 CCAGAGAGAAATCGCTGTGATCATTTGAAGATTAAGAACGGGAGAACTTGCAGCGAGGC 1338
 1021 CCAGAGAGAAATCGCTGTGATCATTTGAAGATTAAGAACGGGAGAACTTGCAGCGAGGC 1080
 1339 AGTGTGAGCTGAGAAACCGCGCAAGCGCTCTTGGAGCAGCAGCCAAAGACAGAG 1398
 1081 AGTGTGAGCTGAGAAACCGCGCAAGCGCTCTTGGAGCAGCAGCCAAAGACAGAG 1140
 1399 CGTGTGCTCAGCTGAGACCGCGCGAGAGAGAGAAAGCGGAGCGCAGAGACAG 1458
 1141 CGTGTGCTCAGCTGAGACCGCGCGAGAGAGAGAAAGCGGAGCGCAGAGACAG 1200
 1459 GAGGCGAAGCGGCGCTGAGCTGAGAGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1518
 1201 GAGGCGAAGCGGCGCTGAGCTGAGAGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
 1519 CAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1578
 1261 CAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
 1579 GAAAGGCGAGCAAACTTGAATGGAAACGAAACCGAGACAGAACTCTGAATCAGAGG 1638
 1321 GAAAGGCGAGCAAACTTGAATGGAAACGAAACCGAGACAGAACTCTGAATCAGAGG 1380
 1639 AACAAGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1698
 1381 AACAAGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
 1699 TTAGAGCTGTGATGACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1758

Db 1441 TTAAAGCTCTGTAATGACAAAGATCAGCTAGAGAAACTTCAGATATCAGGTGT 1500
QY 1759 CCACTGGCAACCCAGAGGCAAGAAATTGAGAGCAGCAAGCTTAGAGAGCTAAGATT 1818
Db 1501 CCACTGGCAACCCAGAGGCAAGAAATTGAGAGCAGCAAGCTTAGAGAGCTAAGATT 1560
QY 1819 GCTGAATCAACCACTTACAGCAGAGTTGCAAGAACTTCAGCAAAATGCTTGGAGACTT 1878
Db 1561 GCTGAATCAACCACTTACAGCAGAGTTGCAAGAACTTCAGCAAAATGCTTGGAGACTT 1620
QY 1879 ATTCCAGAGAAACAGATATCTCAGTACAGCTTAAACCAAGTCCAGCAAGCACTTTGCAT 1938
Db 1621 ATTCCAGAGAAACAGATATCTCAGTACAGCTTAAACCAAGTCCAGCAAGCACTTTGCAT 1680
QY 1939 AGAGACTCGCTTCTTACCTCCAAAGAGCCTTGGAAAGCAAGAGCCTGGCCGAGCAG 1998
Db 1681 AGAGACTCGCTTCTTACCTCCAAAGAGCCTTGGAAAGCAAGAGCCTGGCCGAGCAG 1740
QY 1999 CTCGGGAGAGCTGAGCAGAGTGGAGAGAGAGAGCAGGTCAAAAGCTGCAAGAGATTGAT 2058
Db 1741 CTCGGGAGAGCTGAGCAGAGTGGAGAGAGAGAGCAGGTCAAAAGCTGCAAGAGATTGAT 1800
QY 2059 GTTTTCAACAACCAAGCTGAAAGAACTGAGAGATATCATGCAACAAGCACTCCAGAA 2118
Db 1801 GTTTTCAACAACCAAGCTGAAAGAACTGAGAGATATCATGCAACAAGCACTCCAGAA 1860
QY 2119 CAGAGTCCCTGGAGCAGCGCAGCTGAAAGCAAGAGAGAGAGAGAGCCTGGAG 2178
Db 1861 CAGAGTCCCTGGAGCAGCGCAGCTGAAAGCAAGAGAGAGAGAGAGCCTGGAG 1920
QY 2179 TTAGAGAACAAAGAGAGAGCCTCAGAGACGATTCAAGAAAGGAGCAAGCAATGGCTG 2238
Db 1921 TTAGAGAACAAAGAGAGAGCCTCAGAGACGATTCAAGAAAGGAGCAAGCAATGGCTG 1980
QY 2239 GAGCATGTGAGAGAGAGAGAGCAGCAGCCGCCGGAAGCCCAAGAGAGAGAGAGAG 2298
Db 1981 GAGCATGTGAGAGAGAGAGAGCAGCAGCCGCCGGAAGCCCAAGAGAGAGAGAGAG 2040
QY 2299 AAGAGGAAAGACAGGTGCAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2358
Db 2041 AAGAGGAAAGACAGGTGCAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2100
QY 2359 GACAAACAGAGTGGCTTTTCCATCCGATCAGAGAGCAGCTAAGCTGGCCACCCAGGCA 2418
Db 2101 GACAAACAGAGTGGCTTTTCCATCCGATCAGAGAGCAGCTAAGCTGGCCACCCAGGCA 2160
QY 2419 CCTGGTCTTACACAGAGAAAGGCCGCTTACCATTTCTGACAGAGAGGTGTAAGTGTG 2478
Db 2161 CCTGGTCTTACACAGAGAAAGGCCGCTTACCATTTCTGACAGAGAGGTGTAAGTGTG 2220
QY 2479 GTATATTACGAGGCGGTATACCCCTTTGAAATCCAGAAAGTCAAGATGACATCCAG 2538
Db 2221 GTATATTACGAGGCGGTATACCCCTTTGAAATCCAGAAAGTCAAGATGACATCCAG 2280
QY 2539 CCAGAGATATAGTCAATGTGATGAAAGCAGAGCTGAGAGCCAGATGAGCTTGGAGGA 2598
Db 2281 CCAGAGATATAGTCAATGTGATGAAAGCAGAGCTGAGAGCCAGATGAGCTTGGAGGA 2340
QY 2599 GAGCTGAAAGAGAGAGAGAGGAGTGTTCCTGCAAACTATGACAGAAAGATTCAGAAAT 2658
Db 2341 GAGCTGAAAGAGAGAGAGAGGAGTGTTCCTGCAAACTATGACAGAAAGATTCAGAAAT 2400
QY 2659 GAGGTTCCACTCCAGCAGCAACAGATGACAGATCTGAGCCCTGCCCCCAAGCTG 2718
Db 2401 GAGGTTCCACTCCAGCAGCAACAGATGACAGATCTGAGCCCTGCCCCCAAGCTG 2460
QY 2719 GCTCTGCTGAGAGAGCCCTGCTCTTTGCAAGTACCTCTTGGAGCCTTCAACAACCC 2778
Db 2461 GCTCTGCTGAGAGAGCCCTGCTCTTTGCAAGTACCTCTTGGAGCCTTCAACAACCC 2520
QY 2779 AACAACTGGGAGAGCTTCAAGTTCACAGTGGCCAGAGAGCTCAAAAGAGAGAGAGAG 2838
Db 2521 AACAACTGGGAGAGCTTCAAGTTCACAGTGGCCAGAGAGCTCAAAAGAGAGAGAGAG 2580

QY 2839 GACAACTGGGATACGAGGAGGAGCTGAGCCTTCTGACACCTAGTGTGCTGGCAGTTA 2898
Db 2581 GACAACTGGGATACGAGGAGGAGCTGAGCCTTCTGACACCTAGTGTGCTGGCAGTTA 2640
QY 2899 CGGAGAGATCAGCCTTTACCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2958
Db 2641 CGGAGAGATCAGCCTTTACCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2700
QY 2959 GGCAGAGGTTGAAAGGTGGAAGGCTTACAGGCAAGCCCTGTATTCCTGAGAGCCAA 3018
Db 2701 GGCAGAGGTTGAAAGGTGGAAGGCTTACAGGCAAGCCCTGTATTCCTGAGAGCCAA 2760
QY 3019 AAAAGCAACACCTTAAATTTTAAACAAGTGAAGTATCATCCGTTTGGAAACAGAGAG 3078
Db 2761 AAAAGCAACACCTTAAATTTTAAACAAGTGAAGTATCATCCGTTTGGAAACAGAGAG 2820
QY 3079 ATGTGTGTGTTGAGAAAGTTCAAGGTCAAGAGGTTGTTCCCAAGTCTTACGTGAA 3138
Db 2821 ATGTGTGTGTTGAGAAAGTTCAAGGTCAAGAGGTTGTTCCCAAGTCTTACGTGAA 2880
QY 3139 CTCATTTCCAGGCGCCGTAAAGAAATCCACAGCATCATCTGAGCCTTACTGAAAGTCT 3198
Db 2881 CTCATTTCCAGGCGCCGTAAAGAAATCCACAGCATCATCTGAGCCTTACTGAAAGTCT 2940
QY 3199 GCTAGTCTAAAGAGAGTGGCTTCCCGAGCGCAGAGCCAGCATTCGCGAGAGAGTT 3258
Db 2941 GCTAGTCTAAAGAGAGTGGCTTCCCGAGCGCAGAGCCAGCATTCGCGAGAGAGTT 3000
QY 3259 ATTGCATGTACATACAGAGATTCTGAGCAAGAGATTTAACCTTTGAGCAAGGGGAT 3318
Db 3001 ATTGCATGTACATACAGAGATTCTGAGCAAGAGATTTAACCTTTGAGCAAGGGGAT 3060
QY 3319 GTGATTTGTGTATCCAGAAAGATGTGATCTGTGAGCGGAAACGTTGGCCGACAGTCC 3378
Db 3061 GTGATTTGTGTATCCAGAAAGATGTGATCTGTGAGCGGAAACGTTGGCCGACAGTCC 3120
QY 3379 GGAAGTCTTCCCTTACTATGTGAGGCTTAAAGATTGAGAGGCTTGGAACTGTGGG 3438
Db 3121 GGAAGTCTTCCCTTACTATGTGAGGCTTAAAGATTGAGAGGCTTGGAACTGTGGG 3180
QY 3439 AAAACAGGAGTTTAAAGAAAAAAACCTGAATTTGCCAGGTTATTGCTTCTACCGTGTCT 3498
Db 3181 AAAACAGGAGTTTAAAGAAAAAAACCTGAATTTGCCAGGTTATTGCTTCTACCGTGTCT 3240
QY 3499 ACTGTGCCGAACAATCACTCCTGTGCTCTGGGAGCTGTATTTCTGATCCGAAAAAGAAC 3558
Db 3241 ACTGTGCCGAACAATCACTCCTGTGCTCTGGGAGCTGTATTTCTGATCCGAAAAAGAAC 3300
QY 3559 CAGAGTGAATGTGGAAGAGAGAACTGCAAGCTGCAAGGAAAGAGCCAGATAGGCTGG 3618
Db 3301 CAGAGTGAATGTGGAAGAGAGAACTGCAAGCTGCAAGGAAAGAGCCAGATAGGCTGG 3360
QY 3619 TTTCCAGCAATTTATGTCAAACTTTAAGCCCGGAAACAAGCAAAATCAACCCCACTGAG 3678
Db 3361 TTTCCAGCAATTTATGTCAAACTTTAAGCCCGGAAACAAGCAAAATCAACCCCACTGAG 3420
QY 3679 CTACCCAGAGCCGAGTGCAGCAGCAGAGTGTCCAGGTGATGGGATGTAACAATACCC 3738
Db 3421 CTACCCAGAGCCGAGTGCAGCAGCAGAGTGTCCAGGTGATGGGATGTAACAATACCC 3480
QY 3739 GCCCAAGAGAGAGAGAACTTACCTTACGCAAAAGCCAGATCATCAAGTCTCCAAAG 3798
Db 3481 GCCCAAGAGAGAGAGAACTTACCTTACGCAAAAGCCAGATCATCAAGTCTCCAAAG 3540
QY 3799 GAGGACCCGAGCTGTGTGAAAGAGAGAGTCAATGAGGCTTGTGGCTTTCCATCAAT 3858
Db 3541 GAGGACCCGAGCTGTGTGAAAGAGAGAGTCAATGAGGCTTGTGGCTTTCCATCAAT 3600
QY 3859 TATGTAAAGCTGACACAGACATGAGACCTCAGCCAGCAATGATATGTTGTCCATCC 3918
Db 3601 TATGTAAAGCTGACACAGACATGAGACCTCAGCCAGCAATGATATGTTGTCCATCC 3660

QY	3919	CCCCCTAGGCGTTGAAGTCCCTCAAGAGACCACTATCCATATCACTGCGCCAGAGGA	3919
Db	3661	CCCCCTAGGCGTTGAAGTCCCTCAAGAGACCACTATCCATATCACTGCGCCAGAGGA	3720
QY	3979	TGA 3981	
Db	3721	TGA 3723	
RESULT 2	AF132481	5145 bp	linear
LOCUS	AF132481	Mus musculus	rodent
DEFINITION	AF132481	Mus musculus	rodent
ACCESSION	AF132481.1	GI:4378890	
VERSION	AF132481.1	GI:4378890	
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBLISHED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
FEATURES			
source			
CDS			

BASE COUNT	1448 a	1298 c	1415 g	984 t	
Query Match	71.6%;	Score 3639.4;	DB 10;	Length 5145;	
Best Local Similarity	100.0%;	Pred. No. 0;			
Matches 3640;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;	
QY	259	ATGCTCAGTTTCCCAACACCTTTTGGTGTGATGCTGTGGCCATAACTGTGGAG	318		
DB	1	ATGGCTCAGTTTCCCAACACCTTTTGGTGTGATGCTGTGGCCATAACTGTGGAG	60		
QY	319	GAAGGGCCAAAGCATGACACAGTCCCTTAGCTGAAGCCGATAGCGGGATTATTACT	378		
DB	61	GAAGGGCCAAAGCATGACACAGTCCCTTAGCTGAAGCCGATAGCGGGATTATTACT	120		
QY	379	GGTGATCAAGGGAGAACTTTTTTCCAACTCTGGGTTACCTCAAGCCTGTCTTAGCAAA	438		
DB	121	GGTGATCAAGGGAGAACTTTTTTCCAACTCTGGGTTACCTCAAGCCTGTCTTAGCAAA	180		
QY	439	ATATGGGGCGCTAGCGGACATGAAATAGATGGAAGATGATCAAGTGGAAATTTCCATA	498		
DB	181	ATATGGGGCGCTAGCGGACATGAAATAGATGGAAGATGATCAAGTGGAAATTTCCATA	240		
QY	499	GGCATGAAGCTTATCAACTGAAGCTCAAGGATATAGAGTCCGCTCCACACTTCCCTCT	558		
DB	241	GGCATGAAGCTTATCAACTGAAGCTCAAGGATATAGAGTCCGCTCCACACTTCCCTCT	300		
QY	559	GTCAATGAACAGCAACAGTGGCTATTTCAGTCCACAGCATTTGGTATAGAGGATTT	618		
DB	301	GTCAATGAACAGCAACAGTGGCTATTTCAGTCCACAGCATTTGGTATAGAGGATTT	360		
QY	619	GCTAGCATGCCACCACTCAACAGCTGTGCTCTGTGCCAATGGCTCCATTCCAGTTGTT	678		
DB	361	GCTAGCATGCCACCACTCAACAGCTGTGCTCTGTGCCAATGGCTCCATTCCAGTTGTT	420		
QY	679	GGATGTCTCCACCCCTTAGTACTTCTCTGTCCCTCCAGAGAGAGTCCCTCCCTGGCTAAC	738		
DB	421	GGATGTCTCCACCCCTTAGTACTTCTCTGTCCCTCCAGAGAGAGTCCCTCCCTGGCTAAC	480		
QY	739	GGGGCTCCTCCGCTCATATACAGCCTCTGCTGCTGGTGGCATCTCTGCAGCCATATGGCCA	798		
DB	481	GGGGCTCCTCCGCTCATATACAGCCTCTGCTGCTGGTGGCATCTCTGCAGCCATATGGCCA	540		
QY	799	AAGAATCTTCTTCCAGAGATCTGTGTCCAGGGTCCACAATTAAACCTAAGTTACAGAA	858		
DB	541	AAGAATCTTCTTCCAGAGATCTGTGTCCAGGGTCCACAATTAAACCTAAGTTACAGAA	600		
QY	859	GCACAATCATTCGATGTGCGCAGGGCCCTCCACACACAGAAATGGGCTGTGGCTCAGTCA	918		
DB	601	GCACAATCATTCGATGTGCGCAGGGCCCTCCACACACAGAAATGGGCTGTGGCTCAGTCA	660		
QY	919	TCAAGGCTGAATATACAGGCAAGTTATTCACAAGCAGCACAATAATATAGTGCACATTA	978		
DB	661	TCAAGGCTGAATATACAGGCAAGTTATTCACAAGCAGCACAATAATATAGTGCACATTA	720		
QY	979	ACAGGTCCTCCAGGCAAGAACTATTCTCATGCAATCAAGTTTACCCAGGCTCAGCTGGCT	1038		
DB	721	ACAGGTCCTCCAGGCAAGAACTATTCTCATGCAATCAAGTTTACCCAGGCTCAGCTGGCT	780		
QY	1039	TCAATATGGAATCTTTCTGACATTGATCAAGATGGAAAACTCATGTGCAAGAAATTTATC	1098		
DB	781	TCAATATGGAATCTTTCTGACATTGATCAAGATGGAAAACTCATGTGCAAGAAATTTATC	840		
QY	1099	CTACCTATGACACCAATATGATGTTGGCATGTCTGGTCAGCCATGCGGCCGCTCTGGCT	1158		
DB	841	CTACCTATGACACCAATATGATGTTGGCATGTCTGGTCAGCCATGCGGCCGCTCTGGCT	900		
QY	1159	CCAGAATACATCCCTCTCTCTCCATCAGAAAGATTGCTCCGGCAGTGGGATGTCCGTATA	1218		
DB	901	CCAGAATACATCCCTCTCTCTCCATCAGAAAGATTGCTCCGGCAGTGGGATGTCCGTATA	960		
QY	1219	AGCTCTTCTTCTGTGATCAGAGGCTGCTGAGAGCCGTGTGTAAGAGATGACACGAG	1278		

Db 961 AGCTCTTCTTGTGTGATGAGGCTGCTGAGAGACCCGTCTGAGAGATGAGCAGCAG 1020
Qy 1279 CCAGAGAAAGAACTGCTCTGTGACATTTGAAGATTAAGACGGGAGAACTTCGAGCGAGC 1338
Db 1021 CCAGAGAAAGAACTGCTCTGTGACATTTGAAGATTAAGACGGGAGAACTTCGAGCGAGC 1080
Qy 1339 AGTGTGAGCTGAGAGAGCGCGCCAGCGCTCTTGAGCAGACGCGCAAGAGCAGAG 1398
Db 1081 AGTGTGAGCTGAGAGAGCGCGCCAGCGCTCTTGAGCAGACGCGCAAGAGCAGAG 1140
Qy 1399 CGTGTGCTCAGCTGAGAGCGCGCCGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1458
Db 1141 CGTGTGCTCAGCTGAGAGCGCGCCGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
Qy 1459 GAGGCAAGCGGAGCTGAGAGCTGAGAGACGCTGAGAGAGAGAGAGAGAGAGAGAGAGAG 1518
Db 1201 GAGGCAAGCGGAGCTGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
Qy 1519 CAGCGAGAGAGAGAGAGAGAGAGAGATCGAGAGCGCGCAAAACCGGAGACTG 1578
Db 1261 CAGCGAGAGAGAGAGAGAGAGAGATCGAGAGCGCGCAAAACCGGAGACTG 1320
Qy 1579 GAAAGGACCGGACATTTGAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1638
Db 1321 GAAAGGACCGGACATTTGAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
Qy 1639 AACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1698
Db 1381 AACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
Qy 1699 TTAGAGAGCTCTGATATGACAAAAGCATCAGCTAGAGAGAGAGAGAGAGAGAGAGAG 1758
Db 1441 TTAGAGAGCTCTGATATGACAAAAGCATCAGCTAGAGAGAGAGAGAGAGAGAGAGAG 1500
Qy 1759 CGAGCTGCAACCCAGAGGCAAGAAATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1818
Db 1501 CGAGCTGCAACCCAGAGGCAAGAAATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560
Qy 1819 GCTGAATACCCACTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1878
Db 1561 GCTGAATACCCACTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620
Qy 1879 ATTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1938
Db 1621 ATTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1680
Qy 1939 AAGAGAGCTGCTTCTTACCTCAAAAGAGCTTTGAGAGAGAGAGAGAGAGAGAGAGAG 1998
Db 1681 AAGAGAGCTGCTTCTTACCTCAAAAGAGAGCTTTGAGAGAGAGAGAGAGAGAGAGAG 1740
Qy 1999 CTCGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2058
Db 1741 CTCGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1800
Qy 2059 GTTTTCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2118
Db 1801 GTTTTCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860
Qy 2119 CAGAGAGCTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2178
Db 1861 CAGAGAGCTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1920
Qy 2179 TTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2238
Db 1921 TTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1980
Qy 2239 GAGCATGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2298
Db 1981 GAGCATGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2040
Qy 2299 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2358
Db 2041 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2100

Qy 2359 GACAAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2418
Db 2101 GACAAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2160
Qy 2419 CCTGTGCTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2478
Db 2161 CCTGTGCTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2220
Qy 2479 GTATATTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2538
Db 2221 GTATATTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2280
Qy 2539 CCAGAGAGATATGTATGTATGTATGTATGTATGTATGTATGTATGTATGTATGTATGTAT 2598
Db 2281 CCAGAGAGATATGTATGTATGTATGTATGTATGTATGTATGTATGTATGTATGTATGTAT 2340
Qy 2599 GAGCTGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2658
Db 2341 GAGCTGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2400
Qy 2659 GAGGTTCCCACTCCAGCCAAACAGTGACCGATCTGACATCTGCCCCCTCCCAACTG 2718
Db 2401 GAGGTTCCCACTCCAGCCAAACAGTGACCGATCTGACATCTGCCCCCTCCCAACTG 2460
Qy 2719 GCTCTGCGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2778
Db 2461 GCTCTGCGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2520
Qy 2779 AACCACTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2838
Db 2521 AACCACTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2580
Qy 2839 GACAACTGGGAGATACGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2898
Db 2581 GACAACTGGGAGATACGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2640
Qy 2899 CGGAGAGATCAGAGCTTATACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2958
Db 2641 CGGAGAGATCAGAGCTTATACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2700
Qy 2959 GGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3018
Db 2701 GGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2760
Qy 3019 AAAAGCAACCACTTAATTTTAAACAAAGTACATCAGCTTCTGAGAGAGAGAGAGAG 3078
Db 2761 AAAAGCAACCACTTAATTTTAAACAAAGTACATCAGCTTCTGAGAGAGAGAGAGAG 2820
Qy 3079 ATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3138
Db 2821 ATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2880
Qy 3139 CTCAATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3198
Db 2881 CTCAATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2940
Qy 3199 GCTAGCTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3258
Db 2941 GCTAGCTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3000
Qy 3259 ATTGCAATGTACATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3318
Db 3001 ATTGCAATGTACATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3060
Qy 3319 GTGATTTGTGTGTACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3378
Db 3061 GTGATTTGTGTGTACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3120
Qy 3379 GAGGCTTCCCTTCTAATATGTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3438
Db 3121 GAGGCTTCCCTTCTAATATGTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3180

Qy 3439 AAACAGGAGGATTGAGAAAAAACTGAAATTCGCCAGTTATTCCTTACGCTCT 3438
 Db 3181 AAACAGGAGGATTGAGAAAAAACTGAAATTCGCCAGTTATTCCTTACGCTCT 3240
 Qy 3439 ACTGGTCCCGAACAACCTGAGCTCTGCGGACGCTGATTCGATCCGAAAAAGAAC 3558
 Db 3241 ACTGGTCCCGAACAACCTGAGCTCTGCGGACGCTGATTCGATCCGAAAAAGAAC 3300
 Qy 3559 CCAGGTGATGTTGGGAGAGAACTGCAAGTCCAGGGAAAAAGCGCAGATAGGGTGG 3618
 Db 3301 CCAGGTGATGTTGGGAGAGAACTGCAAGTCCAGGGAAAAAGCGCAGATAGGGTGG 3360
 Qy 3619 TTTCCAGCAAAATTAATGCTAACTTCTAAGCCCGGAAACAAGAAATGACCCCACTGAG 3678
 Db 3361 TTTCCAGCAAAATTAATGCTAACTTCTAAGCCCGGAAACAAGAAATGACCCCACTGAG 3420
 Qy 3679 CTACCAAGACCGCAGTGCAGCAGCAGTGTGCCAGGTGATCGGATGATGATTAACACC 3738
 Db 3421 CTACCAAGACCGCAGTGCAGCAGCAGTGTGCCAGGTGATCGGATGATGATTAACACC 3480
 Qy 3739 GCCCAAGACGATGAGAACTAGCTTACGCAAAAGGCGCAGATGATCAAGTCTCTCAACAG 3798
 Db 3481 GCCCAAGACGATGAGAACTAGCTTACGCAAAAGGCGCAGATGATCAAGTCTCTCAACAG 3540
 Qy 3799 GAGGACCCGAGCTGGTGGAAAGAGAGTCAAGTGGGCAAGTTGGCTCTCCATCCAT 3858
 Db 3541 GAGGACCCGAGCTGGTGGAAAGAGAGTCAAGTGGGCAAGTTGGCTCTCCATCCAT 3600
 Qy 3859 TATGTAAAGCTGACCAAGACATGAGACCCGACCCAGCAATG 3899
 Db 3601 TATGTAAAGCTGACCAAGACATGAGACCCGACCCAGCAATG 3641

RESULT 3
 AF114488 5381 bp mRNA linear PRI 16-JUL-2002

LOCUS Homo sapiens intersecin short isoform (ITSN) mRNA, complete cds.
 AF114488
 AF114488.1 GI:4808824

DEFINITION Homo sapiens (human)
 ACCESSION
 VERSION
 KEYWORDS
 ORGANISM

REFERENCE
 AUTHORS
 TITLE
 JOURNAL

1 (bases 1 to 5381)
 Pucharcos, C., Fuentes, J.J., Pritchard, M. and Estivill, X.
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 Pucharcos, C., Fuentes, J.J., Casas, C., de la Luna, S., Alcantara, S.,
 Arbones, M.L., Soriano, E., Estivill, X. and Pritchard, M.
 Multi-allele cloning of human intersecin (ITSN), a putative
 multivalent binding protein expressed in proliferating and
 differentiating neurons and overexpressed in Down syndrome

Eur. J. Hum. Genet. 7 (6), 704-712 (1999)
 99415290
 MEDLINE
 PUBMED
 10482960

2 (bases 1 to 5381)
 Pucharcos, C., Fuentes, J.J., Pritchard, M. and Estivill, X.
 Direct Submission
 Submitted (16-DEC-1998) Medical and Molecular Genetics Center,
 Cancer Research Institute, V.Hospital de Llo., Avia.
 Castelldefels km. 2,7, Barcelona 08907, Spain

FEATURES
 source
 1..5381
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /chromosome="21"
 /map="21q22.1-q22.2"
 1..5381
 /gene="ITSN"
 /gene="ITSN"
 269..3931
 /gene="ITSN"
 /codon_start=1
 /product="intersecin short isoform"
 /protein_id="AAD29953.1"

gene
 CDS

/db_xref="GI:4808825"
 /translation="MAQPPPPGSGSLDIAITVEERAKHQDFSLKPSKIPGTDGA
 RNFPSGLPQVLAQIWMALDMMNDMMOVSFSTIMKLIKLGQYGLPSLPAM
 KOQPVASAPAFMGSLTAMPPLTANAPPMPSIPVMSPTVSSPVAAPLAPL
 GAPVIOPLPAPFAPPAATLEKSSFSRSGSSQNTLQKAQSFDAVSPVAVEMAVP
 OSSRLKRYQLFNSHDKTMSGLTGPOKRTILTMOSLSLEQALSIWNLSIDIDQKTLA
 EEPFLKHLIDVMSGQPLPVLPEEPIPSFPRVRSQISIVISVSVQRLPEEVL
 LEDEOOLLEKLLPVPFEDKRENERFENLLEKROQLLEQKREORLQLEAEQV
 KREEREOERKQLELEKQLEKORLEERREERREIRREAKREERQLEQLE
 RNRQELNQRNKEQEDIVYLKAKKLELEKREERREERREIRREAKREERQLEQLE
 EISTNSRELRLEITLHLOQLESQMTGRILPEKQILINDQKVOQNSLHRSLSV
 TLKRALKEKLARQLEHLDQLEVEKERSTLQETIDFNQLEKLEIRHNNQQLQKQS
 MEARLKOKEORKEITELKQKEAQRARQERQOMLEHVOEDERHPRKRLPEEVL
 KREESVYKQKEQKQEOADKGLRPHOHEPAKPAVQAPMSTAEKPLTISQENV
 KVVYVRLVYPFESRSHDEITLQPDIVMWGGEWDSQGEPEGLGSELKGTGWAPA
 NYAEKIPENEPVAPVRYVDTSTAPAKLALRETPALATSSSEPSPTPNMADPSST
 WPTSTNKEPPTDMMDAAPSLVTPVAGLQROASATPATATGSSPSVLQGEKYE
 GLQAQALYPMRARKDNLNPNKNDVITVLQODMFWNGEVQKQKMPKSYVXLIISGP
 IRKSTSDMSSESSEPSALIKRVASPAAPVPSGEBFLAMTYSESQGLTFQOQDVL
 VTKKDDPMVTGVDKAGVPSNVVRLKDESGSTAGTSGLSGKPEIAOVIASYAT
 GPEQLTLAPQOLILIRKNPGGMMWEGELQARKKROIQWGPANVYKLISGTSKITPT
 EPRKSTLAALVCYIGNYDTTAQNDDELATNKKOILINLAKEDPMWKEVNGVGLF
 PSNVYKLTMDPSQ"

BASE COUNT 1629 a 1192 c 1304 g 1256 t
 ORIGIN
 Query Match 69.4%; Score 3529.4; DB 9; Length 5381;
 Best Local Similarity 83.7%; Pred. No. 0;
 Matches 4278; Conservative 0; Mismatches 766; Indels 69; Gaps 22;
 Qy 7 GAGGAGAGTGAAGCGGCGGAGGAGGCGGAGCTTGTGCTCCGTAGTACGCGGCT 66
 Db 27 GAGGAGAGTGAAGCGGCGGAGGAGGAGGAGGAGTGTGCTCCGTAGTACGCGGCT 86
 Qy 67 GCGAAGGAGATCCCGAGCGGCTCCGGAACGCGCGGAGGAGGAGGAGCGGCGGCGG 126
 Db 87 GCGAAGGAGATCCCGAGCGGCTCCGGAACGCGCGGAGGAGGAGGAGCGGCGGCGG 135
 Qy 127 GGGATGATGTCGGGCGGCGGAGTCTGGCGCTGCGC-GGGCGTCGGGCTGACCTGA 185
 Db 136 GGGATGATGTCGGGCGGCGGAGTCTGGCGCTGCGCCTCCAGGCGGCGGCTGAGCGGCACTGA 195
 Qy 186 TTTGTGAGAGGAGGCGGCGGCGGAGGCGGCGGAGTGAAGGCTGATCAGCAAGTGA 245
 Db 196 TTTGTGAGAGGAGGCGGCGGAGGCGGAGGCGGAGTGAAGGCTGATCAGCAAGTGA 255
 Qy 246 ACGTAATGAACATGCTGATTTCCCAACCTTTGGTGTGAGCTGCTGATGCTGGGC 305
 Db 256 AAGTAACAGAACATGCTGATTTCCCAACCTTTGGTGTGAGCTGCTGATGCTGGGC 315
 Qy 306 CATACCTGTGAGGAGGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 365
 Db 316 CATACCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 375
 Qy 366 GGGATTTATTAAGTGTGATCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 425
 Db 376 TGAATTCATTAAGTGTGATCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 435
 Qy 426 TGTCTTACACAAATATGAGGCGCTAGCGGAGCATGAATGAGGAGGAGGAGGAGG 485
 Db 436 TGTCTTACACAAATATGAGGCGCTAGCGGAGCATGAATGAGGAGGAGGAGGAGG 495
 Qy 486 GGAATTTTCAATACGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 545
 Db 496 GGAATTTTCAATACGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 555
 Qy 546 CACACTTCCCTGTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 605
 Db 556 TGCACTTCCCTGTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 615
 Qy 606 TATAGAGGAGGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 665
 Db 616 TATAGAGGAGGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 675

QY 666 CATTCAGTTGTGAATGTCTCCACCCCTAGTATCTTGTCCCTCCAGCAGAGGCC 725
 DB 676 CATTCAGTTGTGAATGTCTCCACCCCTAGTATCTTGTCCCTCCAGCAGAGTGCC 735
 QY 726 TCCCTGTGCTAACGGGGCTCTCCCTCATACAGCCTCTGCGCTTGTGCGATCCGTC 785
 DB 736 CCCCCTGGCTAACGGGGCTCCCTGTATACAACTCTGCTGCAATTTGCTCATCCTGC 795
 QY 786 AGCCACATGGCCAAAGAGTCTCTCCAGACATCTGTGTCCAGGGTCAATTTAAACAC 845
 DB 796 AGCCACATGGCCAAAGAGTCTCTCTTATAGATCTGTGTCAGGGTCAACCTAAACAC 855
 QY 846 TAAGTACAGAAAGCAATCATTTGATGTGCGAGGCGCCCTCCAGCAGAGAAATGGG 905
 DB 856 TAATATACAAAGGCAAGTCAATTTGATGTGCGAGTGTCCCAAGTGGCAGAGTGCG 915
 QY 906 TGTGCTCAGTCAATCAAGGCTGAATACAGGCAATTATTCACAGCCACGACAAACTAT 965
 DB 916 TGTTCCTCAGTCAATCAAGGCTGAATACAGGCAATTATTCATATGATGACAAACTAT 975
 QY 966 GAGTGAACCTTACAGGCTGCCAGGCAAACTATTTCTGATGCAATCAAGTTTACCCCA 1025
 DB 976 GAGTGAACCTTACAGGCTGCCAGGCAAACTATTTCTGATGCAATCAAGTTTACCCCA 1035
 QY 1026 GGTCTCAGCTGGCTTCAATATGAAATCTTCTGACATTTGATCAAGATGAAATCTCACTG 1085
 DB 1036 GGTCTCAGCTGGCTTCAATATGAAATCTTCTGACATTTGATCAAGATGAAATCTTCAAG 1095
 QY 1086 AGAAGAAATTTATCTAGCTATGCACTTAATGATGTGCGATGTCTGTGACAGCAGTGC 1145
 DB 1096 AGAAGAAATTTATCTAGCTATGCACTTAATGATGTGCGATGTCTGTGACAGCAGTGC 1155
 QY 1146 GCGCGTCTGCTCCAGAAATACATCCCTCTTCTTCAAGAAAGTTGCTCCGCGATGG 1205
 DB 1156 ACTGTCTGCTCCAGAAATACATCCCTCTTCTTCAAGAAAGTTGATGTGCGAGTGG 1215
 QY 1206 GATGTCCGTCAATAGCTCTTCTTCTGTGATCAGAGGCTGCCGAGAGGCGCTGCACGA 1265
 DB 1216 TATATCTGTCAATAGCTCTTCTTCTGTGATCAGAGGCTGCCGAGAGGCGCTGCACGA 1275
 QY 1266 GATGAGCAGCAGC---CAGAGAAAGAACTGCTGTGACATTTGAAGATTAAGAGCGGA 1322
 DB 1276 AGATGAACAACAACATTTGAAGAAAGAAATTAAGTGAAGTGAAGATTAAGAGCGGA 1335
 QY 1323 GAATCTGAGCGAGGCAAGTGTGAGCTGGAAGCGCCGCGCAAGCGCTCTTGAAGCAGCA 1382
 DB 1336 GAATCTGAGCGGCAAGCTGTGAGCTGGAAGCGAAAGCGCAAGCTCTCTGGAAGCAGCA 1395
 QY 1383 GCGCAAAAGCAGAGCGGCTGCTGAGCTGAGGCGCGCGAGCAGAGAAAGAAAGG 1442
 DB 1396 GCGCAAAAGCAGAGCGGCTGCTGAGCTGAGGCGCGCGAGCAGAGAAAGAAAGG 1455
 QY 1443 GGAAGCGCAGAGCAGAGGCGCAAGCGGAGCTGAGCTGAGAAACAGCTGGAGAAAGA 1502
 DB 1456 TGAAGCGCAGAGCAGAGGCGCAAGCGGAGCTGAGAAACAGCTGGAGAAAGA 1515
 QY 1503 GGGGAGCTGAGCGGCGCAGAGGAGAGAGAGAGAGAGATGAGAGGCGCGAGGC 1562
 DB 1516 GGGGAGCTGAGCGGCGCAGAGGAGAGAGAGAGAGAGAGATGAGAGGCGCGAGGC 1575
 QY 1563 CGCAAAAGCGGAACTGGAAGGCGAGGCAACTTGAATGGAAACGGAACCGAGAGAGA 1622
 DB 1576 TGCAAAACGGGAACCTTGAAGGCGAAGCACTTGAATGGGAACGGAATCGAAGGCAAGA 1635
 QY 1623 ACTCTGAATCAGAGGAGCAAGAGCAGAGGCGACCTGTGCTCTGAAAGGCAAGAGAA 1682
 DB 1636 ACTACTAATCAAAAGAAACAAAGAGCAATAGTTGATGAAAGCAAAAGAAA 1695
 QY 1683 GACTCTGAGTTTGAATGAGAGCTCTGAATGCAAAAGCATCAGCTTGAAGAGAACT 1742
 DB 1696 GACTTTGAGATTTGAATGAGAGCTCTGAATGCAAAAGCATCAGAGAGGAGAACT 1755

QY 1743 TCAGATATCAGGTGCGACTGCGAAACCCAGAGGCGAAGAAATTTGAGAGCAGCAAGATC 1802
 DB 1756 TCAGATATCAGGTGCGACTGCGAAACCCAGAGGCGAAGAAATTTGAGAGCAGCAAGATC 1815
 QY 1803 TAGAGAGTAAAGATTTGTAATCAACCACTTACAGCAGCAGAGTTCAGAAATCTCAGCA 1862
 DB 1816 TAGAGAGTAAAGATTTGTAATCAACCACTTACAGCAGCAGAGTTCAGAAATCTCAGCA 1875
 QY 1863 AATGCTTGAAGACTTATTCAGAGAAACAGATCTCACTGATCCAGTTAAACAGTCCA 1922
 DB 1876 AATGCTTGAAGACTTATTCAGAGAAACAGATCTCACTGATCCAGTTAAACAGTCCA 1935
 QY 1923 GCAGAAACAGTTGCAATAGAGATCCGCTTACCTCAAAAGAGCTTGAAGCAAGAA 1982
 DB 1936 GCAGAAACAGTTGCAATAGAGATCCGCTTACCTCAAAAGAGCTTGAAGCAAGAA 1995
 QY 1983 GCTGGCCCGCAGCAGCTCCGAGAGCAGCTGAGACGAGTGTGAGAGAGAGACAGGTCAA 2042
 DB 1996 ACTAGCTGGCAGCAGCTTACGAGACCACTGAGATGAGTGAAGAAAGAACTAGATCAAA 2055
 QY 2043 GCTGCAAGAGATTTGATTTTCAACCAAGCTGAAGAACTGAGAGATACATACAA 2102
 DB 2056 ACTACAGAGATTTGATTTTCAATATCACTGAAAGAACTAAGAGAAATACAAATTA 2115
 QY 2103 ACAGCAACTCCAGAAAGCAGAGGCTCCGTGAGGCGAGCGGCACTGAAGCAAAAGCAGGA 2162
 DB 2116 GCACCACTCCAGAAAGCAGAGGCTCCGTGAGGCGAGCGGCACTGAAGCAAAAGCAGGA 2175
 QY 2163 GAGGAAGAGCTGAGATTAGAGAAACAAAGAAACGCTCAGAGCAGAGTTCAAGAAAG 2222
 DB 2176 AGGAAGATCATTAGATTTGAAGAAACAAAGAAAGCCCAAGACAGCTCAGAGAAAG 2235
 QY 2223 GGAAGAGCAGTGTGAGAGCATGTGACAGCAGAG---GAGCAGCAGCGCCCGAAGACC 2279
 DB 2236 GGAAGAGCAGTGTGAGAGCATGTGACAGCAGAGCAGAGCATCAGAGCAGCAAGAACT 2295
 QY 2280 CCAGGAGGAGCAGACTGAAAGAGGAGAGACAGTGTGAGAAAGAGGAGCGGAGAGAG 2339
 DB 2296 CCAGGAGGAGGAGGAGGAGGAGAGAGAGTGTGAGAAAGAGAGTGTGAGAGGAGAGAG 2355
 QY 2340 AGCCAGGCGGAAATGCAAGCAAGCAGAGTGTGCTTCCATCCGCACTCAGAGCAGCAGC 2399
 DB 2356 AGCCAGGCGGAAATGCAAGCAAGCAGAGTGTGCTTCCATCCGCACTCAGAGCAGCAGC 2415
 QY 2400 TAACTGCGCACCCAGGCACTCTGTCTTACACAGAGAAAGCGCGCTTACCATTTCTGC 2459
 DB 2416 TAACTGCGCACCCAGGCACTCTGTCTTACACAGAGAAAGCGCGCTTACCATTTCTGC 2475
 QY 2460 ACAGAGAGTGAAGAGTATTAATTAACGAGCGCTGTACCCCTTTGAATCCAGAAATGA 2519
 DB 2476 ACAGAGAGTGAAGAGTATTAATTAACGAGCGCTGTACCCCTTTGAATCCAGAAATGA 2535
 QY 2520 CGATGAGATCAACATCCAGCAGCAGAGATATAGTCAT-----GGTGAATGA 2564
 DB 2536 TGAATGAATCACTATCCAGCAGGAGACATAGTATGATTAAGGAGAAATGAGTGAATGA 2595
 QY 2565 AAGCCAGCTGAGAGCAGAGATGCTTGAAGAGAGAGCTGAAAGGAGAAAGCGAGTGT 2624
 DB 2596 AAGCCAGCTGAGAGCAGAGATGCTTGAAGAGAGAGCTGAAAGGAGAAAGCGAGTGT 2655
 QY 2625 CCTGCAAACTATGAGAGAAAGATTCAGAGAAATGAGTGTCCCACTCCAGCAACCACT 2684
 DB 2656 CCTGCAAACTATGAGAGAAAGATTCAGAGAAATGAGTGTCCCACTCCAGCAACCACT 2715
 QY 2685 GACCATGTGACATCTGCGCCCTGCGCCCAAACTGCGCTGTGCTGAGACCCCTGCTCTTT 2744
 DB 2716 GACCATGTGACATCTGCGCCCTGCGCCCAAACTGCGCTGTGCTGAGACCCCTGCTCTTT 2775
 QY 2745 GCACTGACCTCTTTGAGCCCTCCAGAAACCCCAAACTGAGGAGAGCTTCACTTCACT 2804
 DB 2776 GCACTGACCTCTTTGAGCCCTCCAGAAACCCCAAACTGAGGAGAGCTTCACTTCACT 2835
 QY 2805 GTGGCCAGCAGCTCAAGAGAGAGCAGAAACGAGCAACTGTGGATAGCTGGGCGGCTCA 2864

Db 2836 GTGGCCCAACGACGAAATGAGAAACGAAACGATTAATCTGGATGATGGCAGCCCA 2895
 Qy 2865 GCCTTCTCTGACCGTACTAGTGTGCGCAGTTACGCGAGAGATCAGCTTTTACCAGC 2924
 Db 2886 GCCCTCTCTCAACCGTCCAAAGTCCGCGCAGTTAAAGCAGAGTCCGCTTTACTCCAGC 2955
 Qy 2925 CACAGCCATGCGTCTCTCCCATCTCCGCTCTGCGCCAGAGGTGAAAAGTGAAGGCT 2984
 Db 2956 CACGCGCATGCGTCTCTCCGCTCTCTGCTAGGCGAGGTGAAAAGTGAAGGCT 3015
 Qy 2985 ACAAGCGAAGCCCTGTATCTCCCTGGAGAGCCAAAAAGACAACCACTTTAAATTTTAA 3044
 Db 3016 ACAAGCTCAAGCCCTATATCTTGGAGAGCCAAAAAGACAACCACTTTAAATTTTAA 3075
 Qy 3045 AAGTGAAGTCAATCAGCTTCTGGAAACAGCAAGACATGTGTGTTTGGAGAAAGTTCA 3104
 Db 3076 AAATGATGTCAATCAACGCTCTGGAAACAGCAAGACATGTGTGTTTGGAGAAAGTT 3135
 Qy 3105 TCAGAAAGGTGTGTTCCCAAGTCTTAGTGAACTCAATTCAGGCGCCGTAAAGGAATC 3164
 Db 3136 TCAGAAAGGTGTGTTCCCAAGTCTTAGTGAACTCAATTCAGGCGCCGTAAAGGAATC 3195
 Qy 3165 CACAAGCATCAATCTGAGCCCTACTGAAAGTCTGTAGTCTAAAGAGAGTGGCTTCCC 3224
 Db 3196 TACAAGCATGATTTGTGTTCTTCAGAGAGTCTGTAGTCTAAAGAGAGTGGCTTCCC 3255
 Qy 3225 GCGCGCCAGACCGCATTCCTCCGAGAAAGTTTATTCCTATGACATACATACAGAGTTC 3284
 Db 3256 AGCAGCCCAAGCCGCGCTTTCGGGAGAAAGAAATTTATCCATGTACCTTACAGAGTTC 3315
 Qy 3285 TGAGCAAGGAAATTAACCTTACAGAGAGGAGATGTGTTGTTTACCAAGAAAGATG 3344
 Db 3316 TGAGCAAGGAAATTAACCTTACAGAGAGGAGATGTGTTGTTTACCAAGAAAGATG 3375
 Qy 3345 TGAGTGTGACAGGGAACGTTGCGAGACAGATCCGAGTCTTCCCTTAACTATGTAG 3404
 Db 3376 TGAGTGTGACAGGGAACGTTGCGAGACAGATCCGAGTCTTCCCTTAACTATGTAG 3435
 Qy 3405 GCTTAAAGATTCAGAGGCTCTGGAATCTGTGAGAAACAGGAGTTTAAAGAAAAACC 3464
 Db 3436 GCTTAAAGATTCAGAGGCTCTGGAATCTGTGAGAAACAGGAGTTTAAAGAAAAACC 3495
 Qy 3465 TGAATTTCCAGAGTTATTTGCTTCTAGCGTCTAGTCTCCGGAACATCAACCTGAGC 3524
 Db 3496 TGAATTTCCAGAGTTATTTGCTTCTAGCGTCTAGTCTCCGGAACATCAACCTGAGC 3555
 Qy 3525 TCTTGGGAGCTGATTTCTGATCCGGAAGAAAGAACCCAGTGTGATGTGGGAAGAGACT 3584
 Db 3556 CCTGTGAGCTGATTTTGTATCCGAAAAAGAACCCAGTGTGATGTGGGAAGAGACT 3615
 Qy 3585 GCAAGCTGAGGGAAGAAAGCCGAGATAGGCTGTTTCCAGCAATTAATGTCAAATCTTCT 3644
 Db 3616 GCAAGCAAGTGGGAAGAAAGCCGAGATAGGCTGTTTCCAGCAATTAATGTCAAATCTTCT 3675
 Qy 3645 AAGCCCGGAAACAACAATCAACCCCACTGAGTACCAAGAACCGAGTGCACCCAGC 3704
 Db 3676 AAGCCCGGAGCAACAATCACTCAACAGAGCACCTTAAGTAAACAGATTAAGCGGC 3735
 Qy 3705 AGTGTGCAAGTATCGGAGATGATGATTAACATCCGCGCAAGATGACGATAGCTT 3764
 Db 3736 AGTGTGCAAGTATCGGAGATGATGATTAACATCCGCGCAAGATGACGATAGCTT 3795
 Qy 3765 CAGCAAGGCGCATATCAACGTCCTCAACAAAGAGAGACCCGAGCTGTGGAAAGAGA 3824
 Db 3796 CAACAAGGCGCATATCAACGTCCTCAACAAAGAGAGACCCGAGCTGTGGAAAGAGA 3855
 Qy 3825 AGTGAAGTGGGCAAGTGGGCTTTTCCATCCAAATTAATGAAAGCTGACCAAGATGGA 3884
 Db 3856 AGTGAAGTGGGCAAGTGGGCTTTTCCATCCAAATTAATGAAAGCTGACCAAGATGGA 3915
 Qy 3885 CCCCAGCAGCAATGATCATATGTGTGTCATCCCCCTCAGGCTTGAAGATCTCTCAA 3944

Db 3916 CCAAGCCAGCAATGATCATATGTGTGTCATCCCCCTCAGGCTTGAAGTCTCTCAA 3975
 Qy 3945 GAGACCCATATATCCCATATCACTGCCAGAGAGATGATGGAGATGGAGCTTGAATCATG 4004
 Db 3976 GAGACCCATATATCCCATATCACTGCCAGAGAGATGATGGAGATGGAGCTTGAATCATG 4035
 Qy 4005 TGACTTGCAGATATATCACTTACCTGCTTCTGAGTAAAGAACTCACTGAGAGAGCTTT 4064
 Db 4036 TGACTTGCAGATATATCACTTACCTGCTTCTGAGTAAAGAACTCACTGAGAGAGCTTT 4095
 Qy 4065 ACCTCATTTGACCTTATGTCATGTGATCGAATGTCTGATCACTGCGTGCAGAGCAG 4124
 Db 4096 ACCTCATTTTACCTTATGTCATGTGATCGAATGTCTGATCACTTGCAGAGATGAG 4155
 Qy 4125 AAGC--AAATGGCAATGACAGGAGTGGGCTCTTTTGGGCTTCTCTAGTCACTC 4182
 Db 4156 GAGCAAAATTAACAACACACAGGAGTGGGCTCTTTTGGGCTTCTCTAGTCACTC 4215
 Qy 4183 AGACTGACCGG-CCCGCCTTACACAGGCGCTTCAATAGTTTAAAGATTTTAA 4241
 Db 4216 AAATGACTTTCCTCCACCTTTCGACAGTGTCTTCAATAGTTTAAATTTTAA 4275
 Qy 4242 TGTGATTTTATGCTTTTAAATTAATTCATCAATTAATTTCTTGTGCTATTTGTTT 4301
 Db 4276 TATATATTTATGCTTTTAAATTAATTAATTAATTAATTTCTTGTGCTATTTGTTT 4335
 Qy 4302 TACAAAAACCCCATATCAAGAGTGTCTGTCTGCGAGATTAATAATGCTGTCCGG 4361
 Db 4336 TGCAAAAAGACCCCATATCAAGAGATG----CTGCATGTGCTATTAATAATGTCTCAA 4390
 Qy 4362 CGTACCGTAAATCAGAGCTTGTGTACCTTGTGCGTGTGTCAGTGTCCCAACAT 4421
 Db 4391 ATGTCCATTAATTCGAACATTTGATGTATTTTATTTTGTCCAGTGTACCACTAAT 4450
 Qy 4422 TGTGTAGTTTGGGCTGTCTCC--TCCGTAGACACAGAGAGATGGTGTACTGT 4477
 Db 4451 TGTGCAAGTTTGGGCTTTCCTTCCCTTACATAGAGAGTGCAGAGATTCAGTATCTGT 4510
 Qy 4478 TTTGAAATGTATGTATGATGACGTGACTATGGAAGGCTTATGCTGTCTGTACC 4537
 Db 4511 TTTAAAGACGTATA-GAATGAGCCCAATTAAGGAGAGTGTGTGTGTGTGTGT 4569
 Qy 4538 ATCACTGTATAC--TGTGCGCATGTACATCTGTACGAGAGATGACT--CTTCTCCAT 4595
 Db 4570 ATCACTGTATACCTTGTGATGATGTATACATCTGTACATTAAGAAATTAATGTTTCCA 4629
 Qy 4596 GGTAAACCCACACCGGTATACAGTGTCTCATCTACT--GCATTCATTTTACTTGGAC 4653
 Db 4630 TGGCAAGCTATTAATCTTGTACATGCTGTATATATGATTTAAATTTTATTTGGAC 4689
 Qy 4654 AGTACCTTGTAGCCACTGTAGAGAGAC--CCATGTTCCGTTGTGCTCAGATGACT 4712
 Db 4690 AGTACCTTGTAGCCACTGTAGAGAGACCTGTGTGTGTGTGTGTGTGTGTGTGT 4749
 Qy 4713 AGTGTGCGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4772
 Db 4750 GGTGTAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4809
 Qy 4773 TAAAGCGCAACTGCCAGGCGGTATGATGATGATGATGATGATGATGATGATGATGAT 4827
 Db 4810 GTAGACCAAC-ACCATGAGGTGTTAGATGATGATGATGATGATGATGATGATGATGAT 4868
 Qy 4828 TGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4881
 Db 4869 TGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4928
 Qy 4882 CTTCGTGCAAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4941
 Db 4929 TTTCTTGAAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4987
 Qy 4942 TTTTAACTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5001
 Db 4988 TTTTAACTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5046

Query Match	BASE COUNT	ORIGIN
69.2%	1147 a 1009 c 1123 g 746 t	Score 3517.2, DB 10, Length 4025,

Best Local Similarity 93.4%; Pred. No. 0; Matches 373/4; Conservative 0; Mismatches 243; Indels 21; Gaps 5;									
QY	219	GAGATGAGGCGTCGATCGACGAGGTGAACGTAAATAGAACCATGCGCTCAGTTTCCCAACC	278						
Db	1	GAGATGAGGCGTCGATTATGACAAAGGTAAAGATTAACAGAACCATGCGCTCAGTTTCCGACACC	60						
QY	279	TTTGGGTGTGCGCTGGATGTCTGGGGCATATCTGTGAGAGAAAGGGCCAGCATGACCA	338						
Db	61	TTTGTGTGGAGCGCTGGACATCTGTGGCCATATACCTGTGAGAGAAAGCCAAAGCATGACCA	120						
QY	339	GCAGTTCCTTAGCCCTGAAGCCGATAGCGGGATTATTAATCTGTGATCAAGGAGGAATT	398						
Db	121	GCAGTTCAGAGTTTGAAGCCGATATCTGGAATTAATCACTGTGTATCAAGCAAGGAATT	180						
QY	399	TTTTTTCCAATCTGGGTTACCTGACCGCTGTCTTAGACAATAATATGGCGCTAGCGGACAT	458						
Db	181	CTTTTTCCAATCTGGGTTACCTGACCGCGTCTTAGACAATAATATGGCGCTAGCGGACAT	240						
QY	459	GAATPAAGATGGAAGGATGATCAATGTGGAATTTTCCATAGCCATGAAGCTTATCAAACT	518						
Db	241	GAATPAAGACGGAAGGATGATCAGTGTGGAATTTTCCATAGCCATGAAGCTTATCAAACT	300						
QY	519	GAGGTACAAGGATATAGCTCCCTCCCAACTTCCCTCGTCAATGAAACGACCAACAGT	578						
Db	301	GAGGTACAGAGATATAGCTCCCTCCCGCACTTCCCTCGTCAATGAAAGCAGCAACAGC	360						
QY	579	GGCTAATTTCCAGTCCACCAAGCATTTGGTATAGAGGGAATGTCTAGCATGACCAACTCAC	638						
Db	361	GGCCATCTATGTGTACCAAGCGTTTGGTATAGAGGGAATGTCTGAATGCCACCACTGAC	420						
QY	639	AGCTGTGTCTCTGTGCCAATGGGCTCAATTCAGTTGTGTGAATGTCTCCACCCCTTAGT	698						
Db	421	AGCTGTGTCTCCCTGTGCCAATGGGCTCAATTCAGTTGTGTGAATGTCTCCGCGCTTAGT	480						
QY	699	ATCTTGTGTCCCTCAGAGAGAGTCCCTCCCTGGCTAAAGGGGCTCTCCCGCTCATACA	758						
Db	481	ATCTTGTGTCCCTCAAGACAGAGTCCCTCCCTGGCTAAAGGGGCGCTCTCTGTCAATACA	540						
QY	759	GCCTCTGCGCTGCGTTTGCGCATCTCTGACGACACATGGCGAAAGATTCTTCTCTCAGCAG	818						
Db	541	GCCTCTGCGCTGCAATTTGCTCATCTCTGACGACCAATGTGCCAAAGATTCTTCTCTCAGCAG	600						
QY	819	ATCTGTGTCCAGGATCACAAATTTAAACACTAAGTTACAGAAAGCACAAATCATTCAGTGTGC	878						
Db	601	ATCTGTGTCCAGGATCACAAATTTAAACCTAAGTTACAGAAAGCACAAATCATTCAGTGTGC	660						
QY	879	CAGGCGCCCTTCAGCAGACGAATGGGCTGTGTCTCACTCATCAAGGCTGAAATTCAGGCA	938						
Db	661	CAGGCGCCCTTCAGCGCAGAAATGGGCTGTGTCTCACTCATCAAGGCTGAAATTCAGGCA	720						
QY	939	GTTAATTCAAAGCCAGCAGCAAAAATTATGAGTGGACATTTAAACAGGTCCCAAGCAAGAC	998						
Db	721	GTTAATTCAAAGCTACAGCAAAAGCATAGTGTGGACATTTAAACAGGTCCCAAGCAAGAC	780						
QY	999	TATTTCTATGCAATCAAGTTTATCCCAAGGCTCAGCTGGCTTCATATATGAAATCTTTCTGA	1058						
Db	781	TATTTCTATGCAATCAAGTTTATCCCAAGGCTCAGCTGGCTTCATATGAAATCTTTCTGA	840						
QY	1059	CATTGATCAAGATGGAATACTCACTGCAAGAAATTTATCTTAAGCTATGACCTTAATTGA	1118						
Db	841	CATTGATCAAGATGGAAGCTCACCCGCAAGAAATTTATCTTAAGCTATGACCTTAATTGA	900						
QY	1119	TGTTGCGATGTCTGGTACAGCACTGGCGCGCGTCTGTCGCAAAATCAATCCCTCCCTTC	1178						
Db	901	TGTTGCGATGTCTGTGTACGCACTGGCGCGCGTCTGTCGTCGCAAAATCAATCCCTCCCTTC	960						
QY	1179	CTTACAGAAAGTTGCTCCGCGCAGTGGGAATGTCCGTCATATAGCTCTTCTCTGTGATCA	1238						
Db	961	CTTACAGAAAGTTGCTCTGGCAGTGGGAATGTCCGTCATATAGCTCTTCTCTGTGATCA	1020						
QY	1239	GAGGCTGCTGTGAGAGCCGTGTGTACAGATGAGACAGCCAGAGAAATCTGCTGT	1298						

Db	1021	GC	GC	GCTGCC	CCAG	AGG	CCAT	CGT	CA	GAG	AT	TC	AG	CA	GC	AG	GT	G	AAAA	AAAG	AGT	CGT	CT	108	
Qy	1299	G	A	C	A	T	T	T	G	A	A	G	A	A	A	A	A	A	A	A	A	A	A	A	1358
Db	1081	G	A	C	A	T	T	T	G	A	A	G	A	A	A	A	A	A	A	A	A	A	A	1140	
Qy	1359	C	G	C	C	A	A	G	C	G	C	T	T	T	T	G	A	G	C	A	A	A	A	1418	
Db	1141	C	A	G	G	A	G	C	C	C	C	T	C	T	T	G	A	G	C	A	A	A	A	1200	
Qy	1419	G	C	C	C	A	G	A	G	A	A	A	A	A	A	A	A	A	A	A	A	A	1478		
Db	1201	G	C	C	C	A	G	A	G	A	A	A	A	A	A	A	A	A	A	A	A	A	1260		
Qy	1479	G	C	T	G	A	G	A	A	G	C	T	G	A	A	A	A	A	A	A	A	A	1538		
Db	1261	G	C	T	G	A	G	A	A	G	C	T	G	A	A	A	A	A	A	A	A	A	1320		
Qy	1539	G	A	A	G	A	G	A	T	C	G	A	G	A	G	C	C	G	A	A	A	A	1598		
Db	1321	G	A	A	G	A	G	A	T	C	G	A	G	A	G	C	C	G	A	A	A	A	1380		
Qy	1599	A	T	G	G	A	A	G	G	A	A	C	C	G	A	G	A	C	T	C	T	G	A	1658	
Db	1381	A	T	G	G	A	A	G	G	A	A	C	C	G	A	G	A	C	T	C	T	G	A	1440	
Qy	1659	C	G	T	G	T	C	C	T	G	A	A	G	C	A	A	A	A	A	A	A	A	1718		
Db	1441	C	G	T	G	T	C	C	T	G	A	A	G	C	A	A	A	A	A	A	A	A	1500		
Qy	1719	A	A	A	G	A	T	C	A	G	T	C	A	A	A	A	A	A	A	A	A	A	1778		
Db	1501	A	A	A	G	A	T	C	A	G	T	C	A	A	A	A	A	A	A	A	A	A	1560		
Qy	1779	A	G	A	A	T	T	G	A	G	A	C	A	G	A	A	A	A	A	A	A	A	1838		
Db	1561	A	G	A	A	T	T	G	A	G	A	C	A	G	A	A	A	A	A	A	A	A	1620		
Qy	1839	G	C	A	G	A	G	T	T	G	A	A	A	A	A	A	A	A	A	A	A	A	1898		
Db	1621	G	C	A	G	A	G	T	T	G	A	A	A	A	A	A	A	A	A	A	A	A	1680		
Qy	1899	C	A	G	T	A	C	C	A	T	T	A	A	A	A	A	A	A	A	A	A	A	1958		
Db	1681	C	A	G	T	A	C	C	A	T	T	A	A	A	A	A	A	A	A	A	A	A	1740		
Qy	1959	C	A	A	A	A	G	C	C	T	T	G	A	A	A	A	A	A	A	A	A	A	2018		
Db	1741	C	A	A	A	A	G	C	C	T	T	G	A	A	A	A	A	A	A	A	A	A	1800		
Qy	2019	G	G	T	G	A	G	A	G	A	C	A	G	A	A	A	A	A	A	A	A	A	2078		
Db	1801	G	G	T	G	A	G	A	G	A	C	A	G	A	A	A	A	A	A	A	A	A	1860		
Qy	2079	G	G	A	C	T	G	A	G	A	T	A	C	A	A	A	A	A	A	A	A	A	2138		
Db	1861	G	G	A	C	T	G	A	G	A	T	A	C	A	A	A	A	A	A	A	A	A	1920		
Qy	2139	G	C	A	T	T	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	2198			
Db	1921	G	A																						

QY	2379	CCATCCGGAATGAGGACCGAGCTTAAGCTGCGCAACCCAGGACCCCTGCTCAACAGAGAA	2438
Db	2161	CCATCAATCAAGGCCAGCTTAAGCGG---CCGAGCACCTCGGCCCAAGACAGAA	2217
QY	2439	AGGCCCGCTTACCAATTTCTGCACAGAGAGTGTAAAGTGTATATTAACGAGCGCTGTA	2498
Db	2218	AGGTCGGCTTAATCTCTGCACAGAGAGTGCCAAAGTGTGTATTAACGAGCGCTGTA	2277
QY	2499	CCCCCTTGAATTCAGAAAGTCAGATGAGATCAACATCCAGCCAGAGATATAGTCAT---	2555
Db	2278	CCCCCTTGAATTCAGAAAGTCAGATGAGATCAACATCCAGCCAGAGATATAGTCATAGT	2337
QY	2556	-----GGTGATGAAGCCAGACTGAGAGAGCAGAGATGCTGGAGAGAGACT	2603
Db	2338	TAAAGGGGAATGGGTGATGAAGCCAGACCCGAGAGCCAGAGATGCTGTGAGAGAGAAC	2397
QY	2604	GAAGGGGAAGCAGGATGCTTCCCTGCMAATATAGCAGAAAAATTCAGAAAAATGAGCT	2663
Db	2398	GAAGGGGAAGCAGGATGCTTCCCTGCMAATATAGCAGAAAAATTCAGAAAAATGAGAT	2457
QY	2664	TCCCATCTCAAGCCAAACCGATGACCGATCTGACATCTGCCCCCTGCCCCCAACTGGGCT	2723
Db	2458	TCCCATCTCAAGCCAAACCGATGACCGATCTGACATCTGCCCCCTGCCCCCAACTGGCT	2517
QY	2724	GCGTGAGACCCCTGCTCTTTGCGCAGTGAACCTCTTGTGAGCCCTCCAAACCCCAACAA	2783
Db	2518	GCGTGAGACCCCTGCTCTTTGCGCAGTGAACCTCTTGTGAGCCCTCCAAACCCCAACAA	2577
QY	2784	CTGGGCAGACTTCAAGTTCCAGTGGCCCGACAGACTCAAAAGAGAACGCAAAACGAGCA	2843
Db	2578	CTGGGCAGACTTCAAGTTCCAGTGGCCCGACAGAGCAAAATAGAGAAACGAGAAACGAGCA	2637
QY	2844	CTGGGATACGTGGGCGGCTCAGCCCTTCTCTGACCGTACTGATGTCGGCCAGTTACGGCA	2903
Db	2638	CTGGGATACGTGGGCGGCTCAGCCCTTCTCTGACCGTACTGATGTCGGCCAGTTACGGCA	2697
QY	2904	GAGATCAGCCCTTTACCCCGACAGCCACTGGCTCTCCCAATCTCCCGTCTTGGGCGCA	2963
Db	2698	GAGATCAGCCCTTTACCCCGACAGCCACTGGCTCTCCCAATCTCCCGTCTTGGGCGCA	2757
QY	2964	GGGTGAAAGGTGGAAGGGCTTCAAGCGCAAGCCCTGATATCCCTGGAGAGCCAAAAAGA	3023
Db	2758	GGGTGAAAGGTGGAAGGGCTTCAAGCGCAAGCCCTGATATCCCTGGAGAGCCAAAAAGA	2817
QY	3024	CAACCACTTAATTTTAACAAAAGTGAACGTATCAGCGTCTTGGAACGCAAGACATGTG	3083
Db	2818	CAACCACTTAATTTTAACAAAAGTGAATGTATCAGCGTCTTGGAACGCAAGACATGTG	2877
QY	3084	GTGTTTGGAGAAAGTTCAAGGTCAAGAGGTTGGTTCCCAAGTCTTACGTGAACCTCAT	3143
Db	2878	GTGTTTGGAGAAAGTTCAAGGTCAAGAGGTTGGTTCCCAAGTCTTACGTGAACCTCAT	2937
QY	3144	TTCAAGGGCCGTAAGAGAAATCAACAAGCATGATACAGCCCTTACTGAAATCTCTGCTAG	3203
Db	2938	TTCAAGGGCCGTAAGAGAAATCAACAAGCATGATACAGCCCTTACTGAAATCTCTGCTAG	2997
QY	3204	TCTAAGAGATGGCTTCCCGGCGCGCAAGCCAGCAATTCGCGAGAGAGTTTATTC	3263
Db	2998	TCTAAGAGATGGCTTCCCGGCGCGCAAGCCAGCAATTCGCGAGAGAGTTTATTC	3057
QY	3264	CATGATACATACGAGAGTTCTGAGCAAGAGATTTAACTTTTACGCAAGGGATGTGAT	3323
Db	3058	TATGATACCTTACGAGAGTTCTGAGCAAGAGATTTAACTTTTACGCAAGGGATGTGAT	3117
QY	3324	TGTGTTTACCAAGAAAGATGTGACTGCTGAGACGGGAAACGCTGGGCGCAAGTCCGAGT	3383
Db	3118	TGTGTTTACCAAGAAAGATGTGACTGCTGAGACGGGCAACATGGGCGAGAGTCCGAGT	3177
QY	3384	CTTCCCTTCTACTATGTGAGGCTTAAAGATTCAAGAGGCTCTGGAATCTGCTGGAAAA	3443
Db	3178	CTTCCCTTCTACTATGTGAGGCTTAAAGATTCAAGAGGCTCTGGAATCTGCTGGGAAAA	3237

```

QY 3444 AGGAGTTTGGAAAAAACCTGAAATGTCGCAAGGTTATTGCTCTCTACGCTGCTACTCG 3503
DB 3238 AGGAGTTTGGAAAAAACCTGAAATGTCGCAAGGTTATTGCTCTCTACGCTGCTACTCG 3297
QY 3504 TCCGGAACACATCCCTGCTCTCTGCGGAGCTGATTTGATCCGAAAAAGAACCCAG 3563
DB 3298 TCTTGAACAGCTACCTGCTCTCTGCTGATTTGATCCGAAAAAGAACCCAG 3357
QY 3564 TGGATGTTGGGAAGAGAACTGCAAGCTCGAGGAAAAAGCCGAGATAGGGTGTTC 3623
DB 3358 TGGATGTTGGGAAGAGAACTGCAAGCTCGAGGAAAAAGCCGAGATAGGGTGTTC 3417
QY 3624 AGCAATTTATGCAAACTTTGAAAGCCCGGAAACAAATATCAACCACTGAGCTAC 3683
DB 3418 AGCAATTTATGCAAACTTTGAAAGCCCGGAAACAAATATCAACCACTGAGCTAC 3477
QY 3684 CAAGACCGAGTCCAGCAAGTGTGCGGAGTGTGCGGATTTAGATTACCGCCCA 3743
DB 3478 CAAGACCGAGTCCAGCAAGTGTGCGGAGTGTGCGGATTTAGATTACCGCCCA 3537
QY 3744 GAACGATGACGAATGACCTTTCAGAAAGCCGATCATCAAGCTCTCAAGAGAGGA 3803
DB 3538 GAATGACGATGACCTTTCAGAAAGCCGATCATCAAGCTCTCAAGAGAGGA 3597
QY 3804 CCGGACCTGTTGGAAGAGAACTGAGTGTGCGGCTTCTCCATTCATTAATGT 3863
DB 3598 CCGGACCTGTTGGAAGAGAACTGAGTGTGCGGCTTCTCCATTCATTAATGT 3657
QY 3864 AAAGCTGACCAAGCAATGACCCGAGCCGATTAATGATTTGTCATCCCGCC 3923
DB 3658 GAAGCTGACCAAGCAATGACCCGAGCCGATTAATGATTTGTCAT -CCCCC 3716
QY 3924 TCAGGCTTGAAGTCTCTCAAGAGCCGATTAATGATTTGTCATCCCGAGGAGGAG 3983
DB 3717 TCAGGCTTGAAGTCTCTCAAGAGCCGATTAATGATTTGTCATCCCGAGGAGGAG 3776
QY 3984 GAGATGACGCTTGTATGATGATCTTTCAGATGATCACTTCTGATAGAA 4043
DB 3777 GAGATGACGCTTGTATGATGATCTTTCAGATGATCACTTCTGATAGAA 3836
QY 4044 GAACCTGACGAGAGATTTACCTGATTTGATGATGATGATGATGATGATGATG 4103
DB 3837 GAACCTGACGAGAGATTTACCTGATTTGATGATGATGATGATGATGATGATG 3896
QY 4104 AGTCACTGCTGAGAGAGAGAAATGCAAGCTGCAAGGCTGCTGCTTT 4163
DB 3897 AGTCACTGCTGAGAGAGAAATGCAAGCTGCAAGGCTGCTGCTTT 3955
QY 4164 GGGGCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 4201
DB 3956 -GTGCTTCTGATGATGATGATGATGATGATGATGATGATGATGATG 3992

```

```

RESULT 5
LOCUS AF064243 5287 bp mRNA linear PRI 21-NOV-1998
DEFINITION Homo sapiens interseclin short form mRNA, complete cds.
ACCESSION AF064243
VERSION AF064243.1 GI:3859852
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 5287)
AUTHORS Guipponi, M., Scott, H.S., Chen, H., Schebesta, A., Rossier, C. and
Antonarakis, S.E.
TWO isoforms of a human interseclin (ITSN) protein are produced by
brain-specific alternative splicing in a stop codon
JOURNAL Genomics 53 (3), 369-376 (1998)
MEDLINE 99017974
PUBMED 9799604
REFERENCE 2 (bases 1 to 5287)

```

```

AUTHORS Guipponi, M., Scott, H.S., Chen, H., Schebesta, A., Rossier, C. and
Antonarakis, S.E.
TITLE Direct Submission
JOURNAL Submitted (05-MAY-1998) Genetics and Microbiology, CML, 1 rue
Michel-Servet, Geneva 4 CH-1211, Switzerland
FEATURES
source
1..5287
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="Caxon:9606"
/chromosome="21"
/map="21q22.1-q22.2"
/tissue_type="brain"
/dev_stage="fetus"
107..3769
CDS
/codon_start=1
/product="interseclin short form"
/protein_id="AAC78610.1"
/db_xref="GI:3859853"
/translation="MAORPPRGSLDIWALTVERARHDOQFSLKPISGTIDQAR
RNFPSGLPQVLAQIWLADNDMDVETIAKLILKLOQIQLSALPVM
KOQPAISAPFPMGIAIMPPLTAVAPVWGSPVVGKSPVLSVPTAVPAVAP
GAPVITPLPAPAPATLPKSSPSRPSGSLNTKLOKASFDVASVPAVPAVAP
OSSRLKRYQLFNSHDKTWSGLTGQARTIMOSLPAOLASIMNLSIDIDODGLTA
EEFILAMHLIDVAMSGOPLPVLPEVYIPSEFRVRSSTSVISSTVDORLPEPV
LEBOOOLEKTLPTVPEKRENERPGLLEKROALLBOORKEBLAERAEOR
EKERROERKROLELEKOLEKORLEERBEERKETEREAERLEEROLEME
RNRRELNDORNRKQEDIVLAKKTLFELELANDKQLEGLDORLTLTQRO
EIESTNRKRELAIEITLQOLOESQOMLIPKOILINDOLKOVOONSLHRDLSV
TLKRALFAKELARLDOLDEVEKETSQLEIDIPNOLKELEIRHNKOLOKRS
MEARLEKOEERKILEKORERORAROWLEHMOEDHOKPRLHEERKX
KRESEVKKDGEERKOEADOKRLRFLHQHEPAPKPAVOARWSTREKPLTSAQEN
KVYTRALYPPRESSHDEITTOPEDIVANKEWDESGPGLGGLKGTGFPA
NYAEKIPENEPAPVPTDSTAPAPLALRETPAPLAVTSSESTPNMWABESS
WPTSTNEKPEPNDAMAAQPSLTPVSGOLRORAPATATGSSPSPVVGQEKVE
GLOQALYPMRAKKDNHLENKNDVITVLEQODMMWFGEVQOGKMPKSVYKLSGP
IRKSTSMDSGSESPASLKVYASPAKPVSGEETIAYTESSEQGLTPOQGVIL
VTKDDGMMWGTGVDKAGVPSPNVYRLKDSGSGTAGTSGSGKPEIQAUYASTAT
GPBULTAPGOLILIRKKNPGWMBGLQARCKRQIGMPFANTVKLNLNPGSKTPT
EPKSTALAAVCQVITGMITYDNTAQNDELAFNRKQIINVLNKEDPDMKGEVNGOGLF
PSNVYKLTITDMPDPSQ"
167..406
misc_feature
/note="encodes EH domain"
767..1936
misc_feature
/note="encodes EH domain"
2324..2524
misc_feature
/note="encodes SH3 domain"
2843..3019
misc_feature
/note="encodes SH3 domain"
3110..3286
misc_feature
/note="encodes SH3 domain"
3326..3520
misc_feature
/note="encodes SH3 domain"
3569..3748
misc_feature
/note="encodes SH3 domain"
BASE COUNT 1635 a 1169 c 1227 g 1256 t
ORIGIN
Query Match 67.7%; Score 3440.2; DB 9; Length 5287;
Best Local Similarity 83.8%; Pred. No. 0;
Matches 4154; Conservative 0; Mismatches 743; Indels 58; Gaps 21;

```

345 CCTAGCCTGAAGCCGATAGGGGATTTTACTGCTGATCAAGCGAGAACTTTTTT 404
193 CCAATGTTTAAAGCCAAATATCGAGTCACTTCTGTATCAAGTAAACCTTTTTT 252
405 CCAATCTGGGTACTCTCAGCCCTGTCTTAGCACAAAATATGGCGCTAGCGCATGATTA 464
253 TCAATCTGGGTACTCTCAGCCCTGTCTTAGCACAAAATATGGCGCTAGCGCATGATTA 312
465 CGATGGAAGATGATCAAGTGGAAATTTTCCATAGCCATGAAGCTTATCAACTGAAGCT 524
313 TGAATGGAAGATGATCAAGTGGAAATTTTCCATAGCCATGAAGCTTATCAACTGAAGCT 372
525 ACAAGATATGAGCTCCCTCCACACTTCCCTCTGTCAAGAAACGAAACGAGTGGCTAT 584
373 ACAAGATATGAGCTCCCTCCACACTTCCCTCTGTCAAGAAACGAAACGAGTGGCTAT 432
585 TTCCAGTCCAGCAGATTTGTATAGAGGATGCTAGCATGCGCAGCAGCTCACAGCTGT 644
433 TTCTAGGCGCACACCATTTGGTATGGAGGATAGCGCATGCGCAGCAGCTCACAGCTGT 492
645 TGCTCTGTGCCAATGGGCTCCATTCAGTTGTTGGAATGTCCTCAACCTTAGTATCTTC 704
493 TGCTCAGTGGCCAAATGGGATCCATTCAGTTGTTGGAATGTCCTCAACCTTAGTATCTTC 552
705 TGTCCTCCAGCAGAGTGGCTCCCTGGCTTAAAGGGGCTCCTCCCTCATACAGCCTCT 764
553 TGTTCCACAGAGCTGTGGCCCCCTGGCTTAAAGGGGCTCCTCCCTCATACAGCCTCT 612
765 GCTGCTGTTTGGCATCTCTGAGCCACATAGCCAAAGATTTCTTCCAGCAGATCTG 824
613 GCTGCTGTTTGGCATCTCTGAGCCACATAGCCAAAGATTTCTTCCAGCAGATCTG 672
825 TCCAGGCTCACAATTTAAACATTAATGACAGAGGCAATATTCATTCATTCAGTCCAGCGC 884
673 TCCAGGCTCACAATTTAAACATTAATGACAGAGGCAATATTCATTCATTCAGTCCAGCGC 732
885 CCCTCCAGCAGAGATGGGCTGTGCTCAGTCAATCAAGGCTGAAATACAGGATTTAT 944
733 CCCTCCAGCAGAGATGGGCTGTGCTCAGTCAATCAAGGCTGAAATACAGGATTTAT 792
945 CAACAGCCACGACAAATATAGAGTGAACATTAACAGGCTCCAGCGCAAGAACTATCT 1004
793 CAATAGTCATGACAAATATAGAGTGAACATTAACAGGCTCCAGCGCAAGAACTATCT 852
1005 CATGCAATCAAGTTTAAACCTCAGGCTCAGTGGCTTCAATATGGAATCTTTCTGACATTGA 1064
853 TATGCAATCAAGTTTAAACCTCAGGCTCAGTGGCTTCAATATGGAATCTTTCTGACATTGA 912
1065 TCAAGATGGAATCACTCAGGAGAGAAATTTATCTAGCTATGCACTTAATGATGTTC 1124
913 TCAAGATGGAATCACTCAGGAGAGAAATTTATCTAGCTATGCACTTAATGATGTTC 972
1125 CATGCTGTGTCAGCCACTGCGCCCTGCTGCTCAGAAATACATCCCTCTCTCTGAG 1184
973 TATGCTGTGTCAGCCACTGCGCCCTGCTGCTCAGAAATACATCCCTCTCTCTGAG 1032
1185 AAGAGTTGCTCCGCACTGGGATGTCGTCATTAAGCTCTTCTTGTGATCAGAGCT 1244
1033 AAGAGTTGCTCCGCACTGGGATGTCGTCATTAAGCTCTTCTTGTGATCAGAGCT 1092
1245 GCCTGAGAGCGCTGTGAGAGATGAGCAGAGC---CAGAGAGAAATGCGCTGTAC 1301
1093 ACCAGAGAAACAGTTTAAAGATGAAACAAACAAATTAAGAAATTAACCTGTAC 1152
1302 ATTGGAATGAAGAACGCGAGAACTTGCAGCGAGCAGTGTGAGCTGAGAAAGCGCG 1361
1153 GTTTGAATGAAGAAAGCGAGAACTTTGAAGTGGCAACCTGGAACTGGAGAAACGAG 1212
1362 CCAAGCGCTCTTTGAGCAGCAGCGCAAGAGAGAGAGCGGTTGGCTCAGCTGAGCGCG 1421
1213 GCAAGCTCTCTCGAAGCAGCAGCGCAAGAGAGAGCGCTCGGCCAGCTGAGCGGCG 1272

1422 CGACGAGAGAGAAAGCGCGAGCGCCAGAGACAGAGGCAAGCGGCTGAGAGCT 1481
1273 GAGACGAGAGAGAGAGAGAGCTGTAGCGCCAGAGACAGAGCGCAAAAGCACTGGAAT 1332
1482 GAGAGACAGCTGAGAGAGAGCGGAGCTGAGCGGCAAGCGAGAGAGAGAGAGAGAA 1541
1333 GAGAGACAGCTGAGAGAGAGCGGAGCTGAGCGGCAAGCGAGAGAGAGAGAGAGAGAA 1392
1542 GAGAGATGAGAGAGCGGAGCGGCAAGCGGAACTGAGAAAGCGAGCAACTTGAATG 1601
1393 AGAAATTTGAGAGCGGAGAGCGCTGCAAAACGGGAACTTGAAGCGCAACTTGAATG 1452
1602 GGAACGGAACCGGAGACGAGAACTCTGAATCAGAGGAAACAGAGAGAGAGGCGCT 1661
1453 GGAACGGAATGAGAGAGAGAGAACTTGAATCAGAGGAAACAGAGAGAGAGGCGCT 1512
1662 GGTCTGAGAGCAGAGAGAGAGAGAGCTGTGAGTTGAGTTGAGTTGAGTTGAGTTG 1721
1513 TGTACTGAAAGCAAAAGAAAGAACTTGAATTTGAATTTGAGTTGAGTTGAGTTG 1572
1722 GCATCAGCTAGAGAGAGAACTTCCAGATATCAGTGTGCTGCACTGCGCAACCCAGAGCAAGA 1781
1573 GCATCAACTAGAGAGAGAACTTCCAGATATCAGTGTGCTGCACTGCGCAACCCAGAGCAAGA 1632
1782 AATTGAGAGCAGAAACAGTCTAGAGAGCTAAGAAATGCTGAATCAACCCACTTACAGCA 1841
1633 AATTGAGAGCAGAAACAGTCTAGAGAGCTAAGAAATGCTGAATCAACCCACTTACAGCA 1682
1842 GCAATTCAGAGAACTCAGCAAAATGCTTGAAGAACTTATTCAGAGAAACAGATACCTAG 1901
1693 ACAATTCAGAGAACTCAGCAAAATGCTTGAAGAACTTATTCAGAGAAACAGATACCTAG 1752
1902 TGACCACTTAAACAAAGTCCAGAGAAACAGTTGCTAGAGAGAGCTGCTTCTTACCTCAA 1961
1753 TGACCACTTAAACAAAGTCCAGAGAAACAGTTGCTAGAGAGAGCTGCTTCTTACCTCAA 1812
1962 AAGAGCTTGAAGCAAGAGAGCTGCGCGGAGAGAGCTCAGAGAGAGCTGAGAGAGT 2021
1813 AAGAGCTTGAAGCAAGAGAGCTGCGCGGAGAGAGCTCAGAGAGAGCTGAGAGAGT 1872
2022 GAGAGAGAGAGCAGAGTCAAGGCTGAGAGAGATGATGTTTCAACCAACAGCTGAAGGA 2081
1873 GAGAGAGAGAGCAGAGTCAAGGCTGAGAGAGATGATGATTTCAATATCAGCTGAAGGA 1932
2082 ACTGAGAGAGATATATGCAAGCAAGCACTTCCAGAGCAGAGAGCTCCTGAGAGAGCGCG 2141
1933 ACTGAGAGAGATATATGCAAGCAAGCACTTCCAGAGCAGAGAGCTCCTGAGAGAGCGCG 1992
2142 ACTGAGAGAGATATATGCAAGCAAGCACTTCCAGAGCAGAGAGCTCCTGAGAGAGCGCG 2201
1993 ACTGAGAGAGATATATGCAAGCAAGCACTTCCAGAGCAGAGAGCTCCTGAGAGAGCGCG 2052
2202 TCAGAGAGAGATTCAGAGAAAGGAGCAAGCAATGCTGAGAGATGTCAGCAGAGAG---GA 2258
2053 CCAAGAGAGAGCTCAGAGAAAGGAGCAAGCAGTGGCTGAGAGATGTCAGCAGAGAGAGGA 2112
2259 GCAAGCAGCGCCCGGAGAAACCCACAGAGAGAGCAGCTGAGAGAGAGAGAGAGAGAGCT 2318
2113 GCAATCAGAGAGCAAGAAATCTCCAGAGAGAGAGAACTGAGAAAGGAGAGAGAGTGTCAA 2172
2319 GAAAGAGAGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2378
2173 AAGAGAGAGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2232
2379 CCAATCCGATCAGAGAGCAGCTAAGCTGAGCAACCCAGAGCAACCTGATCTACAGAGAA 2438
2233 CCAATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2292
2439 AGGCGCGCTTACCTTCTGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2498
2293 AGGTCACCTTACCTTCTGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2352
2499 CCCCTTTGAATCCAGAGTCCAGATGAGATCACCATCAGCAGAGAGAGATATGTCAT--- 2555

Db 2353 CCCCTTGAATCAGAAAGCCATGATGAATCACTATCCAGCCAGAGACATATGCTATGCT 2412
 Qy 2556 -----GCTGATGAAAAGCCAGCTGAGAGCCAGATGCTTGGAGAGAGCT 2603
 Db 2413 TAAAGGGGAATGAGTGAATGAAAAGCCAACTGAGAAACCCGGCTGGCTTGGAGAGAAAT 2472
 Qy 2604 GAAAGGAAAGAGCGGATGCTTCCCTGCAAACTATGCAAGAAAGATTCCAGAAAATGAGGT 2663
 Db 2473 AAAAGGAAAGACAGGATGCTTCCCTGCAAACTATGCAAGAAAGATTCCAGAAAATGAGGT 2532
 Qy 2664 TCCCATCTCAGCAAAACAGATGACCGATCTGACATCTGCCCCCTGCCCAAACTGGCTCT 2723
 Db 2533 TCCCATCTCAGTGAAGACAGTACATCTGACATCTGCCCCCTGCCCAAACTGGCTCT 2592
 Qy 2724 GGGTGAAGACCCCTGCTCTTCTTGGCAAGTACCTTCTTGAAGCTCTGACACCCCAACAA 2783
 Db 2593 GGGTGAAGACCCCTGCTCTTGGCAAGTACCTTCTTGAAGCTCTGACACCCCAACAA 2652
 Qy 2784 CTGGGACAGCTTCAGTTCCAGCTGGCCGACGACTCAACGAGAAAGCCAGAAACGGAGCA 2843
 Db 2653 CTGGGACAGCTTCAGTTCCAGCTGGCCGACGACGAGAAAGAAACGAGAAACGGAGTAA 2712
 Qy 2844 CTGGGATACGTGGGCGGCTCAGCTTCTCTGACCGTACCTAGTGGCCAGTTACGAGCA 2903
 Db 2713 CTGGGATGATGGGAGCCGACGCTCTCTCACTGTTCAAGTGGCCGACGAGTTAAGCA 2772
 Qy 2904 GAGATCAGCTTTACCCCAAGCCACAGCCACTGGCTCTCTCCCACTCTCCGCTCTGGGCA 2963
 Db 2773 GAGGTCGCGCTTACCTCCAGCCACGAGCCACTGGCTCTCTCCGCTCTCTGCTGAGGCA 2832
 Qy 2964 GGGTGAAGAGTGAAGGGCTCAAGCGGAGCGGCTGATCCCTGAGAGCCAAAGAAAGA 3023
 Db 2833 GGGTGAAGAGTGAAGGGCTCAAGCGGAGCGGCTGATCCCTGAGAGCCAAAGAAAGA 2892
 Qy 3024 CAACCACTTAATTTTAAACAAAGTACGTCATCACTGTTCTGAAACAGCAAGACATGTG 3083
 Db 2893 CAACCACTTAATTTTAAACAAAGTACGTCATCACTGTTCTGAAACAGCAAGACATGTG 2952
 Qy 3084 GTGGTTTGAAGAACTTCAAGGTCAGAGGGTGTTCCTCCCAAGTCTTACGTAAGTCAAT 3143
 Db 2953 GTGGTTTGAAGAACTTCAAGGTCAGAGGGTGTTCCTCCCAAGTCTTACGTAAGTCAAT 3012
 Qy 3144 TTCAGGGCCGCTAAGGAAATTCAGACATGATCTGGCCCTTACGTAAGTCTGCTAG 3203
 Db 3013 TTCAGGGCCGCTAAGGAAATTCAGACATGATCTGGCCCTTACGTAAGTCTGCTAG 3072
 Qy 3204 TCTAAGAGAGTGGCTTCCCGGCGCCAGCCAGCACTTCCCGGAGAGATTATTCG 3263
 Db 3073 TCTAAGAGAGTGGCTTCCCGGCGCCAGCCAGCCAGCTTTCGGAGAGAAATTATTCG 3132
 Qy 3264 CATGTACATACGAGAGTTCGAGCAGAGAGATTAACTCTTTCAGCAAGGGAGTGTAT 3323
 Db 3133 CATGTACATACGAGAGTTCGAGCAGAGAGATTAACTCTTTCAGCAAGGGAGTGTAT 3192
 Qy 3324 TGTGTTTCCAGAAAGATGAGTCTGCTGAGCAGGAGCGGAGCGGAGAAATTCGGAGT 3383
 Db 3193 TGTGTTTCCAGAAAGATGAGTCTGCTGAGCAGGAGCGGAGCGGAGAAATTCGGAGT 3252
 Qy 3384 CTTCCTTCTAATCTATGAGGCTTAAAGATTCAAGAGGCTCTGCAACTGCTGGGAAAC 3443
 Db 3253 CTTCCTTCTAATCTATGAGGCTTAAAGATTCAAGAGGCTCTGCAACTGCTGGGAAAC 3312
 Qy 3444 AGGAGATTAGGAAAAAACTGAAATTCGCAAGTTATTCCTTCTGACGCTGCTACG 3503
 Db 3313 AGGAGATTAGGAAAAAACTGAAATTCGCAAGTTATTCCTTCTGACGCTGCTACG 3372
 Qy 3504 TCCGAGAACTCACCCCTGGCTCTGGGAGCGTGAATTCGATCCGGAATAAACCAG 3563
 Db 3373 CCCGAGAGCTCACCTGCGCCCTGGTGAAGTATTCGATCCGGAATAAACCAG 3432
 Qy 3564 TGGATGTTGGAAGGAACTGCAAGCTGAGGAGAAAGCCAGATAGGGTGTTC 3623

Db 3433 TGGATGTGGGAAGAGAGCTCAGACAGTGGGAAAAAGCCAGATAGCTGGTTCC 3492
 Qy 3624 AGCAATTAATGTCAACTTCTPAGCCCGGGAAACAAAGCAAAATCACCCTCACTGACTAC 3683
 Db 3493 AGCTAATTAATGTCAACTTCTPAGCCCGGGAAACAAAGCAAAATCACCCTCACTGACTAC 3552
 Qy 3684 CAAGACCGGAGTGCAGCAGAGTGTGCAAGTGTGCGGAGTGAATGATTAACCGGCA 3743
 Db 3553 TPAAGTCAACAGATTAAGCGGAGTGTGCAAGTGTGCGGAGTGAATGATTAACCGGCA 3612
 Qy 3744 GAACGATGACGAATAGCTTCCAGCAAGGCCAGATCACTCAAGTCTTCAACAGAGGA 3803
 Db 3613 GAATGACGATGAGCTGCGCTTCAACAGGCGCAGATCACTCAAGTCTTCAACAGAGGA 3672
 Qy 3804 CCGGAGCTGTGGAAGAGAAAGTCAAGTGGGAGTGTGGGCTTCCATCCATTAATGT 3863
 Db 3673 CCGGAGCTGTGGAAGAGAAAGTCAAGTGGGAGTGTGGGCTTCCATCCATTAATGT 3732
 Qy 3864 AAAGCTGACACAGACATGAGACCCGAGCAGCAATGAATCATATGTTGCCATCCCGCC 3923
 Db 3733 GAAGCTGACACAGACATGAGACCCGAGCAGCAATGAATCATATGTTGCCATCCCGCC 3792
 Qy 3924 TCAGGCTTGAAGTCTTCAAGAGAGCCAGCTATCCATATGACCTGCCAGAGAGATGAT 3983
 Db 3793 TCAGGCTTGAAGTCTTCAAGAGAGCCAGCTATCCATATGACCTGCCAGAGAGATGAT 3852
 Qy 3984 GGAATGCAAGCTTATCATGATGACTTGGACATGATCACTCACTGCTTCACTGATGAA 4043
 Db 3853 GGAATGCAAGCTTATCATGATGACTTGGACATGATCACTCACTGCTTCACTGATGAA 3912
 Qy 4044 GAACCTCAGTGCAGAGAGTTCATTCATTTGACCTTGAAGTGAATGCAAAATGCTG 4103
 Db 3913 GAACCTCAGTGCAGAGAGTTCATTCATTTGACCTTGAAGTGAATGCAAAATGCTG 3972
 Qy 4104 AGTCACTGCTGACAGAGAGAGCA--AATTGCAAGCTGACAGAGGTGAGTCTT 4161
 Db 3973 AGTCACTGCTGACAGAGAGAGCA--AATTGCAAGCTGACAGAGGTGAGTCTT 4032
 Qy 4162 TTTGGGCTTTCTGATCACTGACCTGAC--CGGCGCGCTTCAACAGGCGCTTCAAT 4220
 Db 4033 TTTGGGCTTTCTGATCACTGACCTGACCTTCCCACTTTCAGCAGGCTTTCAT 4092
 Qy 4221 A-GTTTAAAGTATTTTAAATGATTTTAACTTAAATTCATCAAT 4279
 Db 4093 AGTITTAATTAATTTTAAATGATTTTAACTTAAATTCATCAAT 4152
 Qy 4280 ACTTCTTGCCTATTTGTTTAAACAAAGCAACCACTATCAAGAGTGCCTGCTGCG 4339
 Db 4153 GACTTCTTGCCTATTTGTTTAAACAAAGCAACCACTATCAAGAGTGCCTGCTGCG 4207
 Qy 4340 AGGATTAATGCTGTTCCGGGCGTACCGTAACTGAGAGCTTCTGTAACCTTGCCT 4399
 Db 4208 GTGCTATTAATTAATTTTAAATGATTTTAACTTAAATTCATCAAT 4267
 Qy 4400 TGTCCAGTTCCTCAACCAACATGTTGATGTTGGGGCTTCCGCTTCACTGTAAGAG 4455
 Db 4268 TGTCCAGTTCCTCAACCAACATGTTGATGTTGGGGCTTCCGCTTCACTGTAAGAG 4327
 Qy 4456 CAGAGAGATGGGTGACCTGTTTGAATGATGATGATGATGATGATGATGATGATGAT 4515
 Db 4328 CAGAGAGATGGGTGACCTGTTTGAATGATGATGATGATGATGATGATGATGATGAT 4386
 Qy 4516 GGGGTTATGCTTGTGTCGACATGATGATGATGATGATGATGATGATGATGATGAT 4574
 Db 4387 GTGTTTGTGCTGTTGTTGATGATGATGATGATGATGATGATGATGATGATGAT 4446
 Qy 4575 AAGAGTACCT--CTTCCATGATGATGATGATGATGATGATGATGATGATGATGAT 4633
 Db 4447 CATTAAGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4506
 Qy 4634 --GCAATTAATTTTAACTTGAACAGTGAACCTTGAACAGTGAACAGTGAACAGT 4690
 Db 4507 TGCATTAATTTTAACTTGAACAGTGAACAGTGAACAGTGAACAGTGAACAGTGAACAGT 4566

Db 721 GTTATTCAACAGTCACGACAGACCATGAGTGACATTACAGTCCCGACGACAGAAC 780
QY TATTCATGCAATCAAGTTTACCAGGTCAGCTGGCTTCAATTTGAAATCTTTCTTA 1058
Db 999 TATTCATGCAATCAAGTTTACCAGGTCAGCTGGCTTCAATTTGAAATCTTTCTTA 1058
Db 781 TATTCATGCAATCAAGTTTACCAGGTCAGCTGGCTTCAATTTGAAATCTTTCTTA 840
QY 1059 CATTTGCAAGATGAAAACTCAGTCAGAGAAATTTATCTAGTATGCACTTAATTGA 1118
Db 841 CATTTGCAAGATGAAAACTCAGTCAGAGAAATTTATCTAGTATGCACTTAATTGA 900
QY 1119 TGTTCATGCTGTGTGACCACTGCGCCGCTCTGCTTCCAGAAATACATCCCTCTTC 1178
Db 901 TGTTCATGCTGTGTGACCACTGCGCCGCTCTGCTTCCAGAAATACATCCCTCTTC 960
QY 1179 CTTCAGAAAGTTGCTCCGCGAGTGAGATGTCCTGCTTCAAGTCTTCTTCTGTGATCA 1238
Db 961 CTTCAGAAAGTTGCTCCGCGAGTGAGATGTCCTGCTTCAAGTCTTCTTCTGTGATCA 1020
QY 1239 GAGGCTGCTGAGAGCCGTCCTCAGAGATGAGCAGCAGAGAAATGCTGCTGT 1298
Db 1021 GGGGCTGCCCGAGAGCCATCTCAGAGATGAGCAGAGTGGAAGAAAGCTGCTGT 1080
QY 1299 GACATTTGAAGATGAAGCGGAGAACTTCAGCGAGCACTGTGAGCTGAGAGCG 1358
Db 1081 GACATTTGAAGATGAAGCGGAGAACTTCAGCGAGCACTGTGAGCTGAGAGCG 1140
QY 1359 CCGCCAAAGGCTCTTGGAGAGCAGAGCGCAAAGCAGAGCGGTTGGCTTCACTGAGAGG 1418
Db 1141 CAGGCGGCGCTCTCAGAGAGCAGCGCAAGAGCAGAGCGGCTGCTCAGCTGAGAGG 1200
QY 1419 CCGCCAGCAGAGAGAGAAAGCGGAGCGCCAGAGCAGAGAGCGCAAGCGGCTGAG 1478
Db 1201 CCGCCAGCAGAGAGAGAAAGCGGAGCGCCAGAGCAGAGAGCGCAAGCGGCTGAG 1260
QY 1479 GCTGGAAGAGCAGCTGAGAGAGCGGAGCTGAGCGCAGCAGAGAGAGAGAGAG 1538
Db 1261 GCTGGAAGAGCAGCTGAGAGAGCGGAGCTGAGCGCAGCAGAGAGAGAGAGAG 1320
QY 1539 GAAAGGATTCGAGAGCGCGGAGCGGCAAAACCGGAACTGGAAGGCAAGCAGACTTGA 1598
Db 1321 GAAAGGATTCGAGAGCGGAGCGGCAAAACCGGAGCTGGAAGGCAAGCAGACTTGA 1380
QY 1599 ATGAGACCGAAACCGGAGCAGGAACTCTGAATTCAGAGAAACAAGAGAGAGAGGAC 1658
Db 1381 ATGAGACCGAAACCGGAGCAGGAACTCTGAATTCAGAGAAACAAGAGAGAGGAC 1440
QY 1659 CGTGTCTTGAAGGCAAGAGAGAACTCTGAATTTGAATTTGAAGCTCTGAATGACA 1718
Db 1441 CGTGTCTTGAAGGCAAGAGAGAACTCTGAATTTGAATTTGAAGCTCTGAATGACA 1500
QY 1719 AAAAGCTCAGCTTGAAGGAAACTTCAAGATTCAGGTGTCAGCTGGCAACCAAGAGCA 1778
Db 1501 AAAAGCTCAGCTTGAAGGAAACTTCAAGATTCAGGTGTCAGCTGGCAACCAAGAGCA 1560
QY 1779 AGAAATTTGAGAGCAGAACAGTCTAGAGCTAAGAAATTTGCAANTACCCACTTACA 1838
Db 1561 AGAAATTTGAGAGCAGAACAGTCTAAGAGCTGAGAAATTTGCGAAATACCCACTTACA 1620
QY 1839 GCAGCAGTTGCAAGAACTTCAAGAAATGCTTGAAGACTTATTCAGAGAAACAGATCT 1898
Db 1621 GCAGCAGTTGCAAGAACTTCAAGAAATGCTTGAAGACTTATTCAGAGAAACAGATCT 1680
QY 1899 CAGTGCAGGTTTAAACAGTCCAGAGAAACAGTTCATAGAGCTCGCTTCTTACCT 1958
Db 1681 CAGTGCAGGTTTAAACAGTCCAGAGAAACAGTTCATAGAGCTCGCTTCTTACCT 1740
QY 1959 CAAAAGCCTTGAAGCAAAAGAGCTGGCCGCGAGCAGCTCCGAGAGAGCTGAGACA 2018
Db 1741 CAAAAGCCTTGAAGCAAAAGAGCTGGCCGCGAGCAGCTTCCAGAGAGCTGAGACA 1800
QY 2019 GGTGAGAGAGAGACAGGTCAAAGCTGAGAGATTTGTTTCAACAACCAAGCTGAA 2078
Db 1801 GGTGAGAGAGAGACAGGTCAAAGCTGAGAGATTTGTTTCAACAACCAAGCTGAA 1860
QY 2079 GGAATCGAGAGATTAATAGCAACAGCACTCCAGAAAGCAGAGTCCCTGAGAGCAGC 2138
Db 1861 GGAATCGAGAGATTAATAGCAACAGCACTCCAGAAAGCAGAGTCCATGAGCGGAG 1920
QY 2139 GCGACTGAAGCAAGAAAGCAGAGAGAGAGAGAGCTGAGATTAGAGCAAAAGAGAGA 2198
Db 1921 GAGGCTAAAGCAAGAAAGCAGAGAGAGAGAGAGCTGAGATTGAGAGCAAAAGAGAGA 1980
QY 2199 CGCTCAGAGACAGATTCAAGAAAGGAGCAAGCAATGCTGAGCATGTGACAGAGAGA 2258
Db 1981 AGGTCAAGACAGATTCAAGAAAGGAGCAAGCAATGAGAGAGCATGTGACAGAGAGA 2040
QY 2259 GAGAGCGCGCCCGGAAACCCAGAGAGAGAGCAAGCTGAAGAGAGAAAGAGTCTGAG 2318
Db 2041 GAGAGCGCGCCCGGAAACCCAGAGAGAGAGCAAGCTGAAGAGAGAAAGAGTCTGAG 2100
QY 2319 GAAAGAGAGAGCGGAGAGAGAGAGAGAGCCAGCGGAAATGCAAGCAAGAGAGTCTGAG 2378
Db 2101 GAAAGAGAGAGCGGAGAGAGAGAGAGAGCCAGCGGAAATGCAAGCAAGAGAGTCTGAG 2160
QY 2379 CCATCCGATCAGAGAGCCAGCTTACCTGAGCCAGAGCACTTGTCTTACAGAGAA 2438
Db 2161 CCATCCGATCAGAGAGCCAGCTTACCTGAGCCAGAGCACTTGTCTTACAGAGAA 2217
QY 2439 AGGCGCGCTTACCATTTCTGCAAGAGAGTGAATTAAGTGTATTAACCGAGCGCTGTA 2498
Db 2218 AGGCGCGCTTACCATTTCTGCAAGAGAGTGAATTAAGTGTATTAACCGAGCGCTGTA 2277
QY 2499 CCCCTTTGAATCAGAAAGTCAAGATGAGTACCATTCAGAGAGAGATTAAGTAT --- 2555
Db 2278 CCCCTTTGAATCAGAAAGTCAAGATGAGTACCATTCAGAGAGAGATTAAGTAT --- 2337
QY 2556 -----GCTGATGAAGCCAGCTGAGAGAGAGAGAGAGT 2603
Db 2338 TAAAGGAGAGATGGTGAATGAAGCAAGCCAGAGAGCAAGATGGCTTGAAGAGAAC 2397
QY 2604 GAAAGGAGAGAGAGATGGTTCCTGCAAACTATGCAAGAAAGATTCAGAAATGAGGT 2663
Db 2398 GAAAGGAGAGAGAGATGGTTCCTGCAAACTATGCAAGAAAGATTCAGAAATGAGGT 2457
QY 2664 TCCCACTCAGAGCAAAACAGTACAGTACATCTGAGCCCTGAGCCCAAACTGGCTCT 2723
Db 2458 TCCCACTCAGAGCAAAACAGTACAGTACATCTGAGCCCTGAGCCCAAACTGGCTCT 2517
QY 2724 GCGTGAAGCCCTGCTCTTTGCAAGTACCTTTCTGAGCCCTGCAACCCCAACAA 2783
Db 2518 GCGTGAAGCCCTGCTCTTTGCAAGTACCTTTCTGAGCCCTGCAACCCCAACAA 2577
QY 2784 CTGGGCAAGCTTCAAGTTCAGGTGCGCCAGCAGCTCAACAGAGAGAGAGAGAGCA 2843
Db 2578 CTGGGCAAGCTTCAAGTTCAGGTGCGCCAGCAGCAATATGAGAGAGAGAGAGAGCA 2637
QY 2844 CTGGGATAGGTGGGCGGCTCAGCTCTTCTGACCGTACCTAGTGTGGCAGTTTAC 2903
Db 2638 CTGGGATAGGTGGGCGGCTCAGCTCTTCTGACCGTACCTAGTGTGGCAGTTTAC 2697
QY 2904 GAGATCAGCTTTTACCCAGAGCAGAGCACTGGCTCTTCCCATCTCCGCTCTGGGCA 2963
Db 2698 GAGATCAGCTTTTACCCAGAGCAGAGCACTGGCTCTTCCCATCTCCGCTCTGGGCA 2757
QY 2964 GGGTGAAGAGGTGAAGGGCTTCAAGCGCAGCCCTGTATTCCTGAGAGAGAGAGAGAG 3023
Db 2758 GGGTGAAGAGGTGAAGGGCTTCAAGCGCAGCCCTGTATTCCTGAGAGAGAGAGAGAG 2817
QY 3024 CAACCACTTAATTTTAAACAAAGTGAAGTCAAGCTTCTGAGAGAGAGAGAGAGAG 3083
Db 2818 CAACCACTTAATTTTAAACAAAGTGAAGTCAAGCTTCTGAGAGAGAGAGAGAGAG 2877
QY 3084 GTGTTTGAAGAGTTCAAGGTCAAGAGGTTGTTCCCAAGTCTTACGTGAATCTAT 3143
Db 2878 GTGTTTGAAGAGTTCAAGGTCAAGAGGTTGTTCCCAAGTCTTACGTGAATCTAT 2937

3144 TTGAGGCGCCGTAAGAAATCCAGAGATGATCTGAGCCCTAAGTCTGCTAG 3203
 2938 TTGAGGCGCCGTAAGAAATCCAGAGATGATCTGAGCCCTAAGTCTGCTAG 2997
 3204 TCTAAAGAGAGTGGCTTCCCGGCGGCAAGCCAGCCATTCGCCGAGAAAGTTTATTCG 3263
 2998 TCTAAAGAGAGTGGCTTCCCGGCGGCAAGCCAGCCATTCGCCGAGAAAGTTTATTCG 3046
 3264 CATGACATACAGAGAGTCTGAGCAAGAGATTAACTTTGAGCAAGGAGATGTGAT 3323
 3047 ----- 3046
 3324 TGTGTTACCAAGAAAGATGTGATCTGAGCAAGGAGATGTGAT 3383
 3047 ----- 3046
 3384 CTTCCTTTCTAATGATGAGGCTTAAAGATTGAGAGGCTCTGAGACTGCGGAAAC 3443
 3047 ----- 3046
 3444 AGGAGTTTAAAGAAAAAAGTCTGAGCAAGGAGATGTGAT 3503
 3047 ----- 3046
 3504 TCCGCAAGACTCACTGAGCTCTGAGCAAGGAGATGTGAT 3563
 3085 TCTGAAAGAGTCACTGAGCTCTGAGCAAGGAGATGTGAT 3144
 3564 TGGATGTTGAGAGAGAGAGTCAAGGAGATGTGAT 3623
 3145 TGGATGTTGAGAGAGAGAGTCAAGGAGATGTGAT 3204
 3624 AGCAATATATGCAAGTCTGAGCAAGGAGATGTGAT 3683
 3205 AGCAATATATGCAAGTCTGAGCAAGGAGATGTGAT 3264
 3684 CAAGACCGCAGTGCAGCAGAGTGTGAGCAAGGAGATGTGAT 3743
 3265 CAAGACCGCAGTGCAGCAGAGTGTGAGCAAGGAGATGTGAT 3324
 3744 GAAAGATGCAAGTCTGAGCAAGGAGATGTGAT 3803
 3325 GAAAGATGCAAGTCTGAGCAAGGAGATGTGAT 3384
 3804 CCGGAGCTGTTGAGAGAGAGTCAAGGAGATGTGAT 3863
 3385 CCGGAGCTGTTGAGAGAGAGTCAAGGAGATGTGAT 3444
 3864 AAAGCTGACACAGATGAGCCAGCAGCAATGATGATGATGAT 3923
 3445 GAAAGCTGACACAGATGAGCCAGCAGCAATGATGATGATGAT 3503
 3924 TCAAGCTTGAAGAGTCTTCAAGAGAGAGTCAAGGAGATGTGAT 3983
 3504 TCAAGCTTGAAGAGTCTTCAAGAGAGAGTCAAGGAGATGTGAT 3563
 3984 GGAAGTGAAGCTGATGATGATGATGATGATGATGATGAT 4043
 3564 GGAAGTGAAGCTGATGATGATGATGATGATGATGATGAT 3623
 4044 GAACTCACTGAGAGAGATGATGATGATGATGATGATGAT 4103
 3624 GAACTCACTGAGAGAGATGATGATGATGATGATGATGAT 3683
 4104 AGTCACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4163
 3684 AGTCACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3742
 4164 GGGGCTTTCTAAGTCACTGAGAGAGAGAGAGAGAGAGAG 4201
 3743 -GTGCTTTCTAAGTCACTGAGAGAGAGAGAGAGAGAGAG 3779

RESULT 7
 AF114487
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens (human)
 Homo sapiens
 Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 6439)
 Pucharcos, C., Fuentes, J.J., Casas, C., de la Luna, S., Alcantara, S.,
 Athones, M.L., Soriano, E., Estivill, X., and Pritchard, M.
 Alu-aplice cloning of human intersecin (ITSN), a putative
 multivalent binding protein expressed in proliferating and
 differentiating neurons and overexpressed in Down syndrome
 Eur. J. Hum. Genet. 7 (6), 704-712 (1999)
 99415290
 10482960
 2 (bases 1 to 6439)
 Pucharcos, C., Fuentes, J.J., Pritchard, M. and Estivill, X.
 Direct Submission
 Submitted (16-DEC-1998) Medical and Molecular Genetics Center,
 Cancer Research Institute, L'Hospitalet de Llo., Avila.
 Castelldefels km. 2,7, Barcelona 08907, Spain
 location/Qualifiers
 1..6439
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /chromosome="21"
 /map="21q22.1-q22.2"
 1..6439
 /gene="ITSN"
 269..5434
 /gene="ITSN"
 /codon_start=1
 /product="intersecin long isoform"
 /protein_id="AAB29952.1"
 /db_xref="GI:4808823"
 /translation="MAQPPFGSGSDIWAITVEERAKHQDQFSLKPISGFTTGDGA
 RNFPGGGLPVLQAIWALADMDNDGQVSEISIMKIKLQYQVPAALPYM
 KQFVATSSAPAFPMGSIAPMLTAAVAPMGSIPVAGSPYLVSSVAPVPLN
 GAEVLOPLPAPHPAATLPIKSSFSFGSGSOLNTLQKAGSDVAVPVEWAVP
 QSRILKRLPNSHDKTMSHLTPQKRTILMSSSLPQALASLNNISDIDQCKLTA
 EEFILANHLIDVANSQGLPVPVLPETIPSPFRVRSQSIIVISISVQRLPEEFV
 LIEBQDLERKLPVTEFKRKNFERNLEERQALILQKREORLQLEBAE
 RKERERQERKRLLEKQLEKORLEERERREKREIRREKAKERLEORLE
 RNRROELNORKEODIIVLAKKKTLLEFELANDKHLQGLKLODICRLTTO
 EISTNSRELRLEIETHLOOLQESQOMGRILPEKQILNDOLKVOQONSILHDSLV
 TUKRLAEKLEARQHLBDQDEVEKERSTLQGLIDINNOLEKREIRHNQOQKQS
 MEERLQKQERKILEKQKEBAQRAERQDQWLEHQDEHQRPRKLEHEEL
 KRRESYKQDEKQKQADQKRLRLEHQDEBAQRAERQDQWLEHQDEHQRPRKLEHEEL
 KVVYRALYFESRSHDEITOPGDIIVMVGWVDEQTEPGLGELGKGMFPA
 NYAEKIPENVPAPVVKPVTSTAPAPKALRETPALVATSSPPTPNMADFSST
 WPTSNKEPTEIDMDWAALAPSLVPSAQLRORSAFTPATGSSPSVYVULISEP
 GLQKALYFPRAKDNHINRNDVITVLEQDMWMEGELQKGMFPAVYVULISEP
 IRKSTMSDSSSPSLKRVASPAAPVSGEFILMITYESBQDLTFQOQDVL
 VTKKQDMWMTGVDKAGVPSNYVRLKDSGGTAGTKTSLGKKEPIAIVASYTAT
 GPBELLAPOLILIRKKNPGWMEGLQKGRKRGKMPANVYVULISEPSTKTP
 EEPKSTLAIVCOVIGMYDTAQNDDLEAFNKQIIVLNKEDPMKSGVNGVGLF
 PSNYVULITPMDSQKCDLHLDMLPTERKQVITHELIYEEVYVNDLQVTEI
 FQKPLMESELTEKEVAMITVNNKELIMCIIKILKALRYKRSSEGMPPYMGIDTIS
 AQLPHMPYIRFCSRLNGAALIQKTDPADEFEVYKRLAMPKRCGMPSSTLEP
 MORVTRPLIKILIENTPENHDPBLSKALBAELCSQVNEGVAKENSRLEML
 QAHQCGELSEQVFNSTNCLGRKRLHSGKLYKAKSNDELGFLENLDLLOITRK
 PLSSGTDVFPSPKSNLYPQMYKTPILNVLVLYPFGSDGDEFIHSIDIRVYTR
 AESTNERTAVOKIKASSELVITEKKERKAYLVRSQARVIGRLMMVVEGIELD
 CRSHGKSNPYCEVTMSQCHITTKTIQDTLNPKNRSKQFPIRLDEQVLCITYFERQ
 FSPDPLFGRREIRVADIKQDGSKGPVTKLCLLHEVPTGSIIVRLDLQFLDEP"

BASE COUNT
 ORIGIN

1893 a 1523 c 1639 g 1384 t

TITLE Direct Submission
JOURNAL Submitted (05-MAY-1998) Genetics and Microbiology, CMU, 1 rue
Michel-Servet, Geneva 4 CH-1211, Switzerland
FEATURES Location/Qualifiers

source
1. /7247
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.1-q22.2"
/tissue_type="brain"
/dev_stage="fetus"
107..5272
/codon_start=1
/product="intersectin long form"
/protein_id="AAC78611.1"
/db_xref="GI:3859855"
/translation="MAQPTPPFGSLIDIMATVEERAKHDQGFSLKPSIGPTGDOA
RNFPOGSLPQPLAQIMALADNNDGMDVDFSLAMLIKLGLOGYOLPSALPVM
KOOPVASSAPPGMGSIAMPPLTAAVPMGSIPIVMGSPVLSVPLAVPLAN
GAPVIOPLPAFAPAAATLPGSSPSRSGSOLNKLKAKOSFDVASPPVAAAVP
QSRILKTRQLFNSHDKTMSGLTGPARTILMQSSLPQQLASINWLSIDIDQGLTA
EEFLAMHLIDVAMSGQLPVLPPVPIPPSPRRVSGSISVISTSDVQQLPEEPV
LEDEQQLERKLPTVEDKRENERGENLEKRRQALHQQRKEORLEALRABOE
RKREKROERKQLELEKQLEKRELEKREKREKREKREKREKREKREKREKRE
RNRROELNORNKOEIDIVLAKKKKLEPELEALNDKKHOLEGLKODIRCLTQRO
EJESTNKSRELRITAEITTHQQLOESQMLGLRIPKQILNDOLKOVQONSIRHDSLY
TKRALAKELARQHLRDQDEVEKTRSLQSDIDLFNNQLEKLEIRHNKQLOKQKS
MEERLKQKEQERKILEKQKEAQRRAQERDQWLEHVOEDEHQRKQKHEEKL
KEESVKKQDEKQKEQADKGLRPHQOBPAKPAVQAPWSTAEKGLTISAQENL
KVYVYALPFPESRSHDITLQPDIVMGVWDSOTGEFGMLGELKGTGMFPA
NVAEKIPENEPVAPVPTDSTAPAPKALRETAPLAVTSSEPTTPNNMADPEST
WPTSTNEKPTDNDMAQAQPSLTVPASQOLRSPAFTPATAGSSPSVLOGEVE
GIAQALYTPRAKQNDNLNPNKNDIVITVLEQDMMFGEVQGGKGFVYVLSGP
IRKSTMSDGSSESPASLKRVASPAKPVASGEFTAMTYSESEGLTFOGDVIL
VKKQDMDMTGTGDAVGPSPNYVLRKDSGSGTAGTSGSLKKEEIAQVLSYAT
GPEOLTLAAGOLILRRKNMGWEGELARGKROIGWFPAYVILNPGSKITPT
EPKSTALAAVCOVIGWDTYTONDELAPNGQILINVLNKEDPMWKEVNGCVLF
PSNYVLTDMDSQWCDLHLDMLTPERKQGYIHLLIYTBENTYNDLOLVTEI
FQKPLNESELTEKEVAMIFVWKEILMNCIKLKALRVKRWGSEKMPVMDLILS
AQLPHQPIRFSQNLGALIQOKTIDAPKEPKVRLKLEMDPRCKGMLSPILKP
MORVTRVPLIKNILENTPENHPSHLKHALEKABELCSQVGEVREKENSRLKMI
OAHVOEGSEGLVFNSTYNGKPLRKLKELVYKAKNKKELGFLFNDILLITQITK
PLGSGSTKVPSPKSNLYKMYKTPILFNBVLKPLPDSGEPFIHISIDRVYTLR
ASINERTAWQKIKRASELYIETEKKRBKAYLVSQRTGIGRLMNVVSEIEKP
CSHGSKNPYCEVTWGSQCHITKTITQDILNPKWNSQCFIIDLLEVEVCTIVFEQD
FSPDFLGRTEIRVADIKKQDQSGKPVTKCLLHBEVTEIIVRLDLQLFDEP"

misc_feature
167..406
/note="encodes EH domain"
misc_feature
767..1936
/note="encodes EH domain"
misc_feature
2324..2524
/note="encodes EH domain"
misc_feature
2843..3019
/note="encodes SH3 domain"
misc_feature
3110..3286
/note="encodes SH3 domain"
misc_feature
3326..3520
/note="encodes SH3 domain"
misc_feature
3569..3748
/note="encodes SH3 domain"
misc_feature
3836..4390
/note="encodes SH3 domain"
misc_feature
4649..4819
/note="encodes GDS domain"
misc_feature
4895..5143
/note="encodes PH domain"
misc_feature
/note="encodes C2 domain"
BASE COUNT 2156 a 1652 c 1747 g 1692 t
ORIGIN

Query Match 57.0%; Score 2898.4; DB 9; Length 7247;
Best Local Similarity 86.4%; Pred. No. 0;
Matches 3244; Conservative 0; Mismatches 491; Indels 21; Gaps 3;

QY 165 GGGGGGTGGGGGTGCACTGATTTTGTGTAGGGGGGGCCGGCGACCGCCGGAGATG 224
DB 13 GGGGGCGGTAGCGGCACTGATTTTGTGTAGGGGGGGCCGGCGACCGCCGGAGATG 72
QY 225 AGGCGGTGATCAGCAAGTGAACGTAAATGAACCATGCTCATGTTTCCACACCTTTGCG 284
DB 73 AGGCGGTGATCAGCAAGTGAACGTAAATGAACCATGCTCATGTTTCCACACCTTTGCG 132
QY 285 TGGTACCTGATGTCGTGGCCATTAATCTGTGAGGAAAGGGCCAGCATGACCGAGATT 344
DB 133 TGGCAGCTGTGATCTGGGCCATTAATCTGTGAGGAAAGGGCCAGCATGACCGAGATT 192
QY 345 CCTTACCTGAAGCCGATGAGCGGGAATTTATCTGTGTATCAAGGAGGAACCTTTTTT 404
DB 193 CCAATGTTTAAAGCCAAATATCTGATTCATTAATCTGTGTATCAAGGAGGAACCTTTTTT 252
QY 405 CCAATCTGGGTTAAGCTCAGCTGCTTGTAGCAGAAATATGGGCGCTAGCGGACATGAATA 464
DB 253 TCAATCTGGGTTAAGCTCAGCTGCTTGTAGCAGAAATATGGGCGCTAGCGGACATGAATA 312
QY 465 CCAATGAGATGATGATCAAGTGAATTTTCCATAGCAGTAAAGCTTATCAAACTGAAGCT 524
DB 313 TGAATGAAAGATGATCAAGTGAATTTTCCATAGCAGTAAAGCTTATCAAACTGAAGCT 372
QY 525 ACAAGGATATCAGCTCCCTCCCACTCCCTGTCAGAAAGGAAAGCAACGAGTGTAT 584
DB 373 ACAAGGATATCAGCTCCCTCCCACTCCCTGTCAGAAAGGAAAGCAACGAGTGTAT 432
QY 585 TTCCAGTGCAGCAGATTTGTATAGAGAGATGCTAGCATGCCACCACTCAGACTGT 644
DB 433 TTCTAGCGCAGCAGATTTGTATAGAGAGATGCTAGCATGCCACCACTCAGACTGT 492
QY 645 TGTCTCTGTGCATATGGGCTCCATTCAGTGTGGAATGTCTCAACCTTAGATCTTC 704
DB 493 TGTCTCTGTGCATATGGGCTCCATTCAGTGTGGAATGTCTCAACCTTAGATCTTC 552
QY 705 TGTCTCTGTGCAGAGAGGCTCCCTGCTGTAAGGGGGCTCTCCCGTATACAGCTCT 764
DB 553 TGTCTCTGTGCAGAGAGGCTCCCTGCTGTAAGGGGGCTCTCCCGTATACAGCTCT 612
QY 765 GCCTGCTGTGGGCACTCTGTCAGACCATGCGCAAAAGATCTTCTCTGAGAGATCTGG 824
DB 613 GCCTGCTGTGTCTCATCTGTCAGACCATGCGCAAAAGATCTTCTCTGAGAGATCTGG 672
QY 825 TCCAGGGTCACAATTAACCTTAAGTTCAGAGAGGCAACAATCATTCAGATGTGCGCAGCG 884
DB 673 TCCAGGGTCACAATTAACCTTAAGTTCAGAGAGGCAACAATCATTCAGATGTGCGCAGCG 732
QY 885 CCTCTCAGAGAGAGATGGGCTGTGCTCAGATCATCAAGCTGGAATACAGCAGATTA 944
DB 733 CCTCAGAGAGAGATGGGCTGTGCTCAGATCATCAAGCTGGAATACAGCAGATTA 792
QY 945 CAAAGCCAGCAAAATATAGAGAGAGATCACTTAACAGGTCCTCCAGGCAAGAACTATTCT 1004
DB 793 CAAAGCCAGCAAAATATAGAGAGAGATCACTTAACAGGTCCTCCAGGCAAGAACTATTCT 852
QY 1005 CATGCAATCAAGTTTACCAGAGCTCAGCTGCTCAATATAGAAATCTTTCTGACATTGA 1064
DB 853 CATGCAATCAAGTTTACCAGAGCTCAGCTGCTCAATATAGAAATCTTTCTGACATTGA 912
QY 1065 TCAAGATGAAATCACTGTCAGAAAGATTTATCTAGTATGACACTTAATGATGTGCG 1124
DB 913 TCAAGATGAAATCACTGTCAGAAAGATTTATCTAGTATGACACTTAATGATGTGCG 972
QY 1125 CATGCTGTGTCAGCACTGCGCGCTGCTGCTCAGAAATACATCCCTCTTCTTTCAG 1184
DB 973 CATGCTGTGTCAGCACTGCGCGCTGCTGCTCAGAAATACATCCCTCTTCTTTCAG 1032
QY 1185 AAGAGTTGCTCCGCGAGTGGAGTGTGCTCATTAAGCTCTTCTGTCAGATCAAGGCT 1244
DB 1033 AAGAGTTGCTCCGCGAGTGGAGTGTGCTCATTAAGCTCTTCTGTCAGATCAAGGCT 1092

QY	1245	GCCTTAGGAGCGCTGTCAGAGAGATGACGACGAC---CAGAGAGAAACTGCTGTGAC	1301
Db	1093	ACCAAGAGGAAACCAAGTTTATGAAAGTGAACCAACAATTTAGAAAAATTAACCTGTAAAC	1152
QY	1302	ATTTGAAGATTAAGACCGGAGAACTTCGAGGAGGACAGTGTGAGCTGTGAGAACGCGC	1361
Db	1153	GTTTGAAGATTAATAAAACCGGAGAACTTTGAACTGTGGCACTGTGAACTTGTAGAAACGAAG	1212
QY	1362	CCAAACGCTCTTGGAGCAGACGCGCAAAAGACGAGACGCGTTGGCTCACTGTGACGCGC	1421
Db	1213	GCAAGCTCTCTTGGAACAGCAGCGCAAGAGACGAGACGCGCTCGGCCACAGCTGAGCGGAC	1272
QY	1422	CGAGCAGAGAGAGAAAGACCGGAGACCGCCAGAGACAGAGAGCCCAACCGCAGCTGTGACT	1481
Db	1273	GGAGCAGGAGAGAAAGAGACGCGTGAACGCGCCAGAGCAAGCGCAAAAGACCAACTGTGAAC	1332
QY	1482	GGAGAGACGCTGTGAGAAAGCAGACGGGAGCTGTGAGCGGACGACGAGAGAGAGAGAGAA	1541
Db	1333	GGAGAGCAACTGTGAAAAAGCAGCGGAGAGCTTGAACCGCAGAGAGAGAGAGAGAGAGAA	1392
QY	1542	GGAGATCGAGAGCGCGAGGCGGCAAAAACGGGACCTGGAAGGCGGACCAACTTGAATG	1601
Db	1393	AGAAATTTGAGAGGCGAGAGCGCTGCAAAAACGGGACCTTGAAGGCAACGACCAACTGAGTG	1452
QY	1602	GGAAACGGAAACCGGAGACAGGAACCTCTGATATAGAGAAACAAAGACAGAGGCGCACCT	1661
Db	1453	GGAAACGGAAATCCAAAGCAGGAACCTCTAAATCTAAAGAAACAAAGAACAAAGAGACATAGT	1512
QY	1662	GGTCTTGAAGCGAAAGAGAGACACTGTGAGTTGAGTTAGAGCTGTGAATGACAAAAA	1721
Db	1513	TGTACTGAAAGCAAAAGAAAAGAACTTTGGAATTTGAATTTGAAGCTCTTAATGATAAAAA	1572
QY	1722	GCATCAGCTAGAGAGAAACCTTCAGGATATCAGGTGTGACTGTGCAACCCAGAGCGCAGA	1781
Db	1573	GCATCAACTAGAGAGGAAACCTTCAGAGATATCAGATGTGAGTTGACCAACCCAGAGCGCAGA	1632
QY	1782	AATTGAGAGCAGCAACAAGCTTAGAGACTGAAGTTCTGTAATCTCACCTTACCTTACAGA	1841
Db	1633	AATTGAGAGCACAAAACAAATCTAGAGAGATTGGAATTCGCAAAATCACCCACTCTACACCA	1692
QY	1842	GCAGTTGACAGGAATCTCAGCAAAATGCTTGGAGAACTTATTCAGAGAAACAGATCTCAG	1901
Db	1693	ACAAATTCAGGGAATCTCAGCAAAATCTTGGAAAGCTATTCAGAAAAACAGATACCTAA	1752
QY	1902	TGACCAAGTTAAAAACAAGTCTCAGCAACAAGTTGATAGAGACTGTCTTCAACCTCA	1961
Db	1753	TGACCAATTTAAAAACAAGTTTACAGCAABAAGTTTGACAGAGATTCACTGTTTACACTTAA	1812
QY	1962	AAAGACCTTTGGAACCAAAAGAGCTGGCCCGGACAGACTCCGGGAGACCTGTGACGAGGT	2021
Db	1813	AAAGACCTTTAGAGCAAAAAGAACTAGCTCGGACACACTTACGAGACCAACTGTGATGAAGT	1872
QY	2022	GGAGAGAAAGACCAAGTCAAAAGCTGCAGAGATTTGATGTTTTTCAACACAGCTGGAAGA	2081
Db	1873	GGAGAAAGAACTTAAATCAAAACTTACAGAGATTTGATTTTCAATTAATACAGCTGAAGAA	1932
QY	2082	ACTGAGAGAGATACATTAACAAACAGCAACTCCAGAAAGAGAGTCTCTGAGGACGCGC	2141
Db	1933	ACTTAAGAGAAATACCACTATTAAGCAACAATCTCCAGAACCAAAAGTCAATGTAGAGCTGAAC	1992
QY	2142	ACTGAAGCAGAAAGACAGAGAGAGAGAGCTGTGAGTTTGAAGACCAAAAGAAAGACGC	2201
Db	1993	ACTGAAACAGAAAGAAACAAAGACGAAGATCTATTAATTTAGAAAAACAAAAGAAAGAGC	2052
QY	2202	TCAGAGACGAGTTCAAGAAAGGGAACAGCAATGCTGTGAGCACTGTGCACGACGAGG--GA	2258
Db	2053	CCAAAGAGAGAGCTCAGGAAAGGGAACACAGTGGCTGTGAGCATGTGCACGACGAGAGACGA	2112
QY	2259	GCAGGCAACGCCCCCGGAAACCCCAAGAGAGGACAGACTGAAGAGAGGAAAGCAAGTGTGAG	2318
Db	2113	GCATTAAGAACCAAGAAAACTTCAAGAAAGGAAAAACCTGAAAGGGAGAGAGTGTCA	2172
QY	2319	GAAAGAGAGGCGGAAGAGAGAGCCAGACCGGAATGTCAAGACAGACGAGATCGGCTTTT	2378

Db	2173	AAAGAAGATGCGAGAGAAAAAGCCAAACAGAGAGCAACAGACAAAGCTGGGTCCGCTTTT	2232
OY	2379	CCATCCGATCAGAGAGCCAGCTAAAGCTGCGCCACCCAGGCACTCTGGTCTACACAGAGAA	2438
Db	2223	CCATCAACCCAAAGAACAGAGCTAAGCCAGCGTGTCCAGGCACTCGTGTCACTGCAGAAAA	2292
OY	2439	AGGCCCCGTTACCAATTTCTGCAACAGAGAGTGTAAAAGTGATATTAACGAGCGCTGT	2498
Db	2293	AGGTCCACTTACATTTCTGCAACAGAAAAATGTAAAAAGTGATATTAACCGGCACTGT	2352
OY	2499	CCCCTTTGAATCCAGAAAGTCACAGTGCATGCATCCATCCAGCCAGAGATATATGTCAT---	2555
Db	2353	CCCCTTTGAATCCAGAAAGCCATGTATGAATCTACTATCCAGCCAGAGACATATGTCATGCT	2412
OY	2556	-----GGTGATGAAACCACAGACTGAGAGCCAGAGATGGCTTGAGAGAGAGCT	2603
Db	2413	TAAAGGGGAATGGGTGGATGAATAACCAAACTGAGAGAACCCGGCTGGCTTGAGAGAGATTT	2472
OY	2604	GAAAGGAGAGAGCGGATGCTTCCCTGCACAACTATGACAGAAAAAGATTCCAGAAAAATGAGT	2663
Db	2473	AAAAGGAAAGACAGAGGTGGTTCCTTCGCAAACTATGACAGAGAAAAATCCAGAAAAATGAGT	2532
OY	2664	TCCCACTTCAGCCAAACAGTACCCGATCTGACATCTGGCCCTGCCCCCAAACTGGCTCT	2723
Db	2533	TCCCGCTCCAGTGAACCAAGTACGTATTCACAACTGCCCCCTGCCCCCAAACTGGGCTT	2592
OY	2724	GGGTGAGAGCCCTCTCTCTTTGGCCAGAGACCTCTTCTGAGGCCCTCCAAACCCCCCAAA	2783
Db	2593	GGGTGAGAGCCCCCGCCCTTTGGGAGATTAACCTCTTCAGAGGCCCTCCAGACCCCTTAATTA	2652
OY	2784	CTGGGCAAGCTTCAGTTTCAAGTGGCCACAGAGCTCAACGAGAGGCCAGAAACGAGCAA	2843
Db	2653	CTGGGCGAGCTTCAGCTCCAGGTGGCCCAACAGCAAGAAATGAGAAACAGAAACGAGTAA	2712
OY	2844	CTGGGATACGTGGGCGGCTCAGCCTTCTCTGACCTGATACGTAGCTGGCCAGTTACGGCA	2903
Db	2713	CTGGGATGCAATGGGCAAGCCCAAGCCCTCTCTCAACGTTCCAAAGTGCACGGCCAGTTAAAGCA	2772
OY	2904	GAGATCAAGCCCTTTTACCAGGCAACAGGCACTGGCTCTCCCCCAATCTCCGTCCTGGGCA	2963
Db	2773	GAGGTCCGCTTTTACTCCAGGCAAGGCACTGGCTCTCCCCCGTCTCTGTGCTAGGCCA	2832
OY	2964	GGGTGAAAGGTGGAGGGGCTACAGGCGCAAGCCCTGTATCCCTGAGAGCCAAAAAGA	3023
Db	2833	GGGTGAAAGGTGGAGGGGCTACAGAGCTCAAGCCCTATATCTTTGAGAGCCAAAAAGA	2892
OY	3024	CAACCACTTAATTTTAAACAAAGTGCATCTACCGTCTGTGGAACAGCAACATGTG	3083
Db	2893	CAACCACTTAATTTTAAACAAAGTGTATCTACCGTCTGTGGAACAGCAACATGTG	2952
OY	3084	GTGGTTTGGAGAAATTCAGAGGTTCAGAGGGGTGGTTCCCAAGTCTTACGTAACCAATCAT	3143
Db	2953	GTGGTTTGGAGAAATTCAGAGGTTCAGAGGGGTGGTTCCCAAGTCTTACGTAACCAATCAT	3012
OY	3144	TTCAAGGCGCGTAAAGAAATCCAAACATCGATCTAGTGCCTTACTGAAAGTCTCTGTAG	3203
Db	3013	TTCAAGGCGCCATAAGAAAGTCTAACAGCATGGAATCTTGTTCTTCAAGAGTCTCTGTAG	3072
OY	3204	TCATAAGAGAGTGGCTTCCCGGCGCCAGAGCCAGCAATTCCTGGAGAGAGTTTATTC	3263
Db	3073	TCATAAGCGAGTGGCTTCTCCAGAGCCAGAGCCGAGTGGTTCCGGAGAAATAATTAATTC	3132
OY	3264	CATCTAACATACAGAGGTTCTGAGGAAAGAGATTTTAACTTTTCAGAAAGGGATGTGAT	3323
Db	3133	CATCTAACATTCACAGAGTTCTGAGCAAGAGATTTTAACTTTTCAGAAAGGGATGTGAT	3192
OY	3324	TGTGTTTACCAAGAAAGATGGTGACTGTGTGACGGGAACGGTGGGCGACAAGTCCGAGT	3383
Db	3193	TTTGTGTTTACCAAGAAAGATGGTGACTGTGTGACGGGAACAGTGGGCGACAAGTCCGAGT	3252
OY	3384	CTTCCCTTCTAATATGTAGAGCTTAAAGATTCAGAGGGCTTGTGAATCTGCTGGAAAAAC	3443

Db 290 TCACATGTTAAAGCAATATCTGGAATTCACTACTGATCAAGTAGAAATCTTTT 349
 Qy 404 TCACATCTGGGTTACCTCAAGCTGCTTACGACAAATATGGGCGGTACGGGACATGATA 463
 Db 350 TTCAATCTGGGTTACCTCAAGCTGCTTACGACAAATATGGGCGGTACGGGACATGATA 409
 Qy 464 AGAGTGAAGAGTGAATGAGTGAATTTTCCATGAGCCATGAGGCTTAACTGAAGC 523
 Db 410 ATGATGAAGAGTGAATGAGTGAATTTTCCATGAGCCATGAGGCTTAACTGAAGC 469
 Qy 524 TACAAGATATCAAGCTCCCTCCACACTTCCCTGTGATGAAACAGCAACAGTGGCTA 583
 Db 470 TACAAGATATCAAGCTCCCTCCACACTTCCCTGTGATGAAACAGCAACAGTGGCTA 529
 Qy 584 TTTCCAGTGCACACGATTTTGTATAGAGAGGATTTGCTAGCATGACCACTCAGAGCTG 643
 Db 530 TTTCCAGTGCACACGATTTTGTATAGAGAGGATTTGCTAGCATGACCACTCAGAGCTG 589
 Qy 644 TTGCTCTGTGTCATGAGGCTCATTTCCAGTTGTTGATGATGTCACACCTTATGATCTT 703
 Db 590 TTGCTCTGTGTCATGAGGCTCATTTCCAGTTGTTGATGATGTCACACCTTATGATCTT 649
 Qy 704 CTGCTCTGTGTCACAGAGTCCCTCCCTGCTGCTAAAGGAGCTTCCCTGCTATACAGCTC 763
 Db 650 CTGCTCTGTGTCACAGAGTCCCTCCCTGCTGCTAAAGGAGCTTCCCTGCTATACAGCTC 709
 Qy 764 TTGCTCTGTGTCAGCATCTCTGAGCCACATGGCCAAAGATTTCTTCTGACAGATCTG 823
 Db 710 TTGCTCTGTGTCAGCATCTCTGAGCCACATGGCCAAAGATTTCTTCTGAGTATCTG 769
 Qy 824 GTCCAGGCTCAATTAACACTAAGTTACAGAGGACCAATCATTCGATGCGCAGC 883
 Db 770 GTCCAGGCTCAATTAACACTAAGTTACAGAGGACCAATCATTCGATGCGCAGC 829
 Qy 884 CCCCTCCAGCAGAGATGGGCTGTGCTCACTCATCAAGGCTGAATACAGGACATTAT 943
 Db 830 TCCCAACAGTGCAGAGTGGGCTGTGCTCACTCATCAAGGCTGAATACAGGACATTAT 889
 Qy 944 TCAACAGCAGCAGAAACTATGATGAGCATTTAAACAGGTCCTCCAGGACAGAACTATTC 1003
 Db 890 TCAATAGTCATGACAAACTATGATGAGCATTTAAACAGGTCCTCCAGGACAGAACTATTC 949
 Qy 1004 TCATGCAATCAAGTTTACCCAGGCTCAGCTGCTTCAATGAACTCTTCTGACATTG 1063
 Db 950 TTATGCAATCAAGTTTACCCAGGCTCAGCTGCTTCAATGAACTCTTCTGACATTG 1009
 Qy 1064 ATCAAGATGAAAACTCACTGCAAGAAATTTATCTGATGATGACCTTAATGATGTTG 1123
 Db 1010 ATCAAGATGAAAACTTACAGCAGAGAAATTTATCTGATGATGACCTTAATGATGTTG 1069
 Qy 1124 CCATGCTGTGTCAGCATCTGCGCCCTGCTGCTTCCAGAAATACATCCCTCTTCTTCA 1183
 Db 1070 CCATGCTGTGTCAGCATCTGCGCCCTGCTGCTTCCAGAAATACATCCCTCTTCTTCA 1129
 Qy 1184 GAAAGATTCGCTCCGCGAGTGGATGTCGCTATAGGCTCTTCTGATGATGAGGC 1243
 Db 1130 GAAAGATTCGCTCCGCGAGTGGATGTCGCTATAGGCTCTTCTGATGATGAGGC 1189
 Qy 1244 TGCCTGAGAGCCCTGCTCAGAGATGAGCAGCAGC--CAGAGAGAACTGCTGTGA 1300
 Db 1190 TACCAAGAGAACCAATTTAGAAAGTGAACAACAATTAAGAAAGAAATTAACCTGTAA 1249
 Qy 1301 CATTTGAAGATTAAGAAAGCGGAGAACTTGAAGCGGACAGTGTGAGCTGAGAAAGCCG 1360
 Db 1250 CATTGAAGATTAAGAAAGCGGAGAACTTGAAGCGGACAGTGTGAGCTGAGAAAGCCG 1309
 Qy 1361 GCCAAGCGCTTTGGAGAGCAGCGGAGAAAGAGAGAGAGGCTTGGCTCAGCTGAGCGG 1420
 Db 1310 GCCAAGCGCTTTGGAGAGCAGCGGAGAAAGAGAGAGAGGCTTGGCTCAGCTGAGCGG 1369
 Qy 1421 CCGAGCAGAGAGAGAAAGAGCGGAGAGCGCAGAGAGAGAGAGCGCAGCTGAGC 1480
 Db 1370 CCGAGCAGAGAGAGAAAGAGCGTGAAGCGCAGAGAGAGAGAGCGCAAAAGCAACTGGAAC 1429

Qy 1481 TGGAGAGCAGCTGGAGAGCAGCGGAGCTGGAGCGGACGAGAGAGAGAGAGAG 1540
 Db 1430 TGGAGAGCAGCTGGAGAGCAGCGGAGCTGGAGCGGACGAGAGAGAGAGAGAG 1489
 Qy 1541 AGAGATCAGAGAGCGGAGCGGAGAGCGGAACTGGAAGCGAGCAGCAACTTGAAT 1600
 Db 1490 AAGAAATTTAGAGAGCGGAGCGGAGCTGCAAAACGGGAACTTGAAGAGCAACCTTGA 1549
 Qy 1601 GGGAGCGGAAACCGAGACAGAGAACTCTGAATCAGAGAAACAAGAGCAGAGAGCGC 1660
 Db 1550 GGGAGCGGAACTCAGAGCGGAGAACTTGAATCAGAGAAACAAGAGCAGAGAGCTAG 1609
 Qy 1661 TGGCTGGAAGGAGAGAGAGAGCTGGAAGTTTGAATGAGAGCTCTGAATGACAAA 1720
 Db 1610 TTGCTGGAAGGAGAGAGAGAGAGCTTGAATGAGAGCTCTGAATGATTAATA 1669
 Qy 1721 AGCATCAGTGAAGAGAGAACTTCAAGATATCAAGTGTGATGCACTGCAACCCAGAGC 1780
 Db 1670 AGCATCAACTGAAGAGAGAACTTCAAGATATCAAGTGTGATGCAACCCAGAGC 1729
 Qy 1781 AAATTGAGAGCAGAGCAAGTCTGAGAGCTGAGAAATTCGTAATCAACCACTTAAGC 1840
 Db 1730 AAATTGAGAGCAGAGCAAGTCTGAGAGCTGAGAAATTCGTAATCAACCACTTAAGC 1789
 Qy 1841 AGCAGTTGAGAGAGATCTCAGAGAAATGCTTGAAGAGCTTATCCAGAGAAACAATCTCA 1900
 Db 1790 AACATTAACAGAGATCTCAGAGAAATGCTTGAAGAGCTTATCCAGAGAAACAATCTCA 1849
 Qy 1901 GTGACAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1960
 Db 1850 ATGACCAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1909
 Qy 1961 AAAAGCCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2020
 Db 1910 AAAAGCCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1969
 Qy 2021 TGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2080
 Db 1970 TGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2029
 Qy 2081 AACTGAGAGAGATACATAGCAACAGCAACTCCAGAGAGAGAGAGAGAGAGAGAGAGAG 2140
 Db 2030 AACTGAGAGAGATACATAGCAACAGCAACTCCAGAGAGAGAGAGAGAGAGAGAGAGAG 2089
 Qy 2141 GACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2182
 Db 2090 GACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2131

RESULT 11
 AK074554
 LOCUS 2131 bp mRNA linear PRI 03-SEP-2002
 DEFINITION Homo sapiens cDNA FLJ90073 f18, clone HEMBA1004110, highly similar to Homo sapiens interseclin short form mRNA.
 ACCESSION AK074554
 VERSION AK074554.1 GI:22760070
 KEYWORDS oligo capping; f18 (full insert sequence).
 SOURCE Homo sapiens
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 Isegai,T., Ota,T., Nishikawa,T., Hayashi,K., Otsuki,T., Sugiyama,T., Suzuki,Y., Nagai,K., Sugano,S., Ishii,S., Kawai-Hio,Y., Saito,K., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Kojima,S., Nagahara,K., Masuno,Y., Ono,T., Okano,K., Yoshikawa,Y., Aotsuka,S., Sasaki,N., Hattori,A., Okumura,K., Iwayanagi,T. and Ninomiya,K.
 TITLE NED0 human cDNA sequencing project
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 2131)
 AUTHORS Isegai,T. and Otsuki,T.

TITLE Direct Submission
JOURNAL Submitted (25-MAR-2002) Takao Isogai, Helix Research Institute,
Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
(E-mail: isogai@helix.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing:
Research Association for Biotechnology; cDNA library construction:
Institute of Medical Science, University of Tokyo, Laboratory of
Genome Structure, Human Genome Center; cDNA 5'- & 3'-end one pass
sequencing and clone selection: Helix Research Institute (supported
by Japan Key Technology Center etc.).

FEATURES
source
1. 2131
/organism="Homo sapiens"
/mol_type="rRNA"
/db_xref="taxon:9606"
/clone="HEMBA1004110"
/issue_type="whole embryo, mainly head"
/clone_lib="HEMBA1"
/dev_stage="embryo, 10 weeks"
/note="cloning vector: pME18SFL3"

BASE COUNT 694 a 469 c 326 g 442 t
ORIGIN

Query Match 31.7%; Score 1610; DB 9; Length 2131;
Best Local Similarity 85.5%; Pred. No. 0;
Matches 1892; Conservative 0; Mismatches 295; Indels 15; Gaps 3;

QY 45 GTTGTCTCGTAGTACGCGCGCTCGCAAGGAGATCCGAGCGGCTCCGGACGCGCG 104
DB 1 GTTGTCTCGTAGTACGCGCGCTCGCAAGGAGATCCGAGCGGCTCCGGACGCGA--- 57
QY 105 GAGCAGGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 164
DB 58 -----CAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 109
QY 165 -CGCGCGCTCGCGCGCTCGCATGATTTGTGTGAGGGGCGCGCGCGCGCGCGCGCGAGAT 223
DB 110 AGCGCGCGCTGAGCGCGCGCATGATTTGTCTCTGGGCGCGCGCGCGCGCGCGCGAGAT 169
QY 224 GAGCGCTGATGAGCAAGGTGAACGTATAGAACCATGCGCTCACTTTCCACACTTTTCG 283
DB 170 GAGCGCTGATGAGCAAGGTGAACGTATAGAACCATGCGCTCACTTTCCACACTTTTCG 229
QY 284 GTGTGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 343
DB 230 GTGTGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 289
QY 344 TCTTACCTGAGAGCGGATAGCGGATTTATTACTGTGATCAAGCGAGGAACTTTTTC 403
DB 290 TCTTACCTGAGAGCGGATAGCGGATTTATTACTGTGATCAAGCGGAACTTTTTC 349
QY 404 TCCAACTCTGGGTTACCTCAAGCTGTCTTACCAAAATATGCGCGCTAGCGGAGATGATA 463
DB 350 TCCAACTCTGGGTTACCTCAAGCTGTCTTACCAAAATATGCGCGCTAGCGGAGATGATA 409
QY 464 AGCATGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 523
DB 410 AGCATGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 469
QY 524 TACAAGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 583
DB 470 TACAAGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 529
QY 584 TTTTCCAGTCAAGCATTTGTGATAGAGGATTTCTAGCATGAGCACTACAGCTG 643
DB 530 TTTTCCAGTCAAGCATTTGTGATAGAGGATTTCTAGCATGAGCACTACAGCTG 589
QY 644 TTGCTCTGTGCGCAATGCGCTCATTCAGTTGTGGAATGTCTCAACCTTATGATCTT 703
DB 590 TTGCTCTGTGCGCAATGCGCTCATTCAGTTGTGGAATGTCTCAACCTTATGATCTT 649
QY 704 CTGTCTCTCAAGAGAGTGTCTCCCTGTGCTAACGCGGCTCTCCCGTATACAGCTTC 763

DB 650 CTGTCTCTCAAGAGAGTGTCTCCCTGTGCTAACGCGGCTCTCCCGTATACAGCTTC 709
QY 764 TGTCTGTGTTGGCATCTCTGACGCAATGCGCAAGATTTCTTCTTACAGATCTG 823
DB 710 TGTCTGTGTTGGCATCTCTGACGCAATGCGCAAGATTTCTTCTTACAGATCTG 769
QY 824 GTTCCAGGTCAGAAATTAACATAAGTTACAGAGCAATTCATTCATTCGCGCAGG 883
DB 770 GTTCCAGGTCAGAAATTAACATAAGTTACAGAGCAATTCATTCATTCGCGCAGG 829
QY 884 CCCCTCCAGCAGCAAGATGAGCTGTGCTCAGTCAATCAAGCTGAAATACAGGAGTTAT 943
DB 830 TCCACAGTGTGCAAGATGAGCTGTGCTCAGTCAATCAAGCTGAAATACAGGAGTTAT 889
QY 944 TCAACAGCAGCAAGAACTATAGTGAACACTTAACAGTCTCCAGGCAAGAACTATTC 1003
DB 890 TCAATATGTCATGACAAACTATAGTGAACACTTAACAGTCTCCAGGCAAGAACTATTC 949
QY 1004 TCAATGATCAAGTTTACCCAGGCTCAGCTGCTTCAATATGAAATCTTCTGACATTC 1063
DB 950 TTAATGATCAAGTTTACCCAGGCTCAGCTGCTTCAATATGAAATCTTCTGACATTC 1009
QY 1064 ATCAAGATGAGAAATCACTGACAGAAATTTATCTTACCTATGACCTAATGATGTTG 1123
DB 1010 ATCAAGATGAGAAATCACTGACAGAGAAATTTATCTTACCTATGACCTAATGATG 1069
QY 1124 CCATGCTGTGTCAGCACTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1183
DB 1070 CTATGCTGTGTCAGCACTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1129
QY 1184 GAAAGTTGCTGCTGCGCAATGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1243
DB 1130 GAAAGTTGCTGCTGCGCAATGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1189
QY 1244 TGCTGTGAGAGCGCTGCTGACAGATGAGCAGCAGC---CAGAGAGAAATCGCTGTGA 1300
DB 1190 TACAGAGAGAACCACTTATGAGATGAGCAACCAATTAAGAAATTAATCTGTGA 1249
QY 1301 CATTTGAGATTAAGAGCGGAGAACTTGAGCGAGCGAGTGTGAGCTGAGAAACGCGC 1360
DB 1250 CGTTTGAAGATTAAGAGCGGAGAACTTGAGCGAGCGAGTGTGAGCTGAGAAACGAA 1309
QY 1361 GCCAAGCGCTCTTGTGAGCAGCAGCGCAAGAGCAGAGAGCGGTTGCTCAGCTGAGCGCG 1420
DB 1310 GCCAAGCGCTCTTGTGAGCAGCAGCGCAAGAGCAGAGAGCGGTTGCTCAGCTGAGCGCG 1369
QY 1421 CCGAGCAGAGAGAGAAAGAGCGGAGCGCAGAGCAGAGAGAGAGCGCAGCTGAGAGC 1480
DB 1370 CCGAGCAGAGAGAGAAAGAGCGGAGCGCAGAGCAGAGAGAGAGCGCAGCTGAGAGC 1429
QY 1481 TGGAGAGAGCTGTGAGAGAGCGGAGCTGAGCGGAGCGGAGAGAGAGAGAGAGAGAGAG 1540
DB 1430 TGGAGAGAGAGCTGTGAGAGAGCGGAGCTGAGCGGAGCGGAGAGAGAGAGAGAGAGAG 1489
QY 1541 AGAGATGAGAGAGCGGAGCGGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1600
DB 1490 AGAGATGAGAGAGCGGAGCGGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1549
QY 1601 GGGAGCAGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1660
DB 1550 GGGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1609
QY 1661 TGTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1720
DB 1610 TGTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1669
QY 1721 AGCATCAGTGAAGAGAAATCTTCAAGATATAGAGTGTGAGCTGAGCAACCCAGAGAGAG 1780
DB 1670 AGCATCAGTGAAGAGAAATCTTCAAGATATAGAGTGTGAGCTGAGCAACCCAGAGAGAG 1729
QY 1781 AAATGAGAGCAGAGAAAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1840

1146	GC	CGCCGCTCGCTCCAGAAATCATCCCTTCCTTCCTTCAGAAAGTTGCTCCGGCAGTGG	1205
1123	AC	CTGCTCGCTCCAGAAATCATTCCTTCCTTCCTTCAGAAAGTTGCTTCGATCGCATGTGG	1182
1206	GA	TGCTCCGCTCATTAAGCTCTTCTCTGTGAATCAGAGGCTGCCTGAGAGCCGTCGTCA	1265
1183	TA	TATCTCTGTCAATAGCTCAACATCTGTATGATCAGAGGCTTACCAAGAGAACCGTTTNGA	1242
1266	GG	AATGAGAGAGCAGC---CAGAGAGAAATCTGCTGTGAATTTTGAAAGTTAAGACGGGGA	1322
1243	AG	ATGAATCAACAAACAATTGAAAGAAATTTACCTGTAACTGTTTGAAGTTAAGAAAGCGGGA	1302
1323	GA	ATCTTCAGCGAGGACAGTGTGAGCTGGAAGAGCGCCGCAAGCGCTCTTGAGCAGCA	1382
1303	GA	ACTTTTGAAGTGGCAACCTGGAATCTGGAAGAAAGAAAGCAAGCTCTCTGTGAACAGCA	1362
1383	GC	CGCAAAAGCAGAGAGCGGTTGGCTCAGCTGAGCGCGCCGAGCAGAGAGAGAAAGACG	1442
1363	GC	CGCAAGAGAGCAGAGCGCCCTGGCCCACTGGAAGCGGCGGAGCAGAGAGAAAGACG	1422
1443	GG	AGCGCCAGAGAGCAGAGCGCCAGCGGACCTGGAAGCTTGAAGAAAGCACTGTGAGAGCA	1502
1423	TG	AGCGCCAGAGAGCAGAGCGCCAAAGCAACTGGAATCTGGAAGAAAGCAACTGTGAAGCA	1482
1503	GC	GGGAGACTGAGCGGCAAGAGAGAGAGAGAGAGAGAGATCAGAGAGCGCGAGGC	1562
1483	GC	GGGAGCTTAAACAGCAGAGAGAGAGAGAGAGAGAGAGAGAAATTTGAGAGCGCAGAGGC	1542
1563	CG	CAAAACCGGAACTGGAAGCGCAGCGCAACTTGAATGAGAAACGGAACCGGAGCAGGA	1622
1543	TG	CAAAACCGGAACTTGAAGGCAACGCAACTTGAATGAGGAGACGGAATCGAAGGCAAGA	1602
1623	AC	CTCTGAATCAGAGGAAACAAGAGCAGAGAGGCAACCTGCTCTTGAAGGCAAGAGGA	1682
1603	AC	CTCAATCAATCAAGAAACAAGAAACAAGAGACATTACTTGTATCGAAAGCAAAAGAAAA	1662
1683	GA	CTCTGAGATTGAGTAGAAGCTCTGAATGACAAAAAGCATCAGTAGAAGAGAAACT	1742
1663	GA	CTTTGGAATTTGAATTTGAAGCTCTTAATGATTAAGAAAGCATCACTAGAAGGAAACT	1722
1743	TC	AGATATCAGGTGCGATGCGCAACCCAGAGCGAAGAAATTTGAGCAGCAACAAGTC	1802
1723	TC	AGATATCAGATGTGCGATTGACCAACCAAGGCAAGAAATTTGAGCAGCAACAATTC	1782
1803	TA	GAGAGCTAAGAAATTCGTGAATACCCCACTTACAGAGCAGTTGCAAGAAATCTCAGCA	1862
1783	TA	GAGAGTTGGAATTTGCGGAATCACCACTTACAGAGCAACATTTACAGAAATCTCAGCA	1842
1863	AA	ATCTTGAAGACTTATTTCCAGAGAAACAGATACCTGAGCCAGTTAAACAAGTCA	1922
1843	AA	ATCTTGAAGACTTATTTCCAGAGAAACAGATACCTGAGCCAGTTAAACAAGTCA	1902
1923	GC	AGAAACAGTTGCAATAGAGACTGCTTCTTACCTCAAAAGCGCTTGAAGCAAGGA	1982
1903	GC	AGAAACAGTTTGCACAGAGATTCCTTGTTAACCTTAAAGCGCTTGAAGCAAGGA	1962
1983	GC	TGGCCCGGAGAGCGTCCGGGAGCACTGAGCGAAGTTGAGAGAGACCAAGTCAAA	2042
1963	AC	TAGCTCGGAGCACTTACAGAGACCAACTGATGAATTTGAGAAAGAACTTAGTCAAA	2022
2043	GC	TGAGAGAGATTGATGTTTTCAACAAACAGACTGAAGAACTGAGAGATACATAGCAA	2102
2023	AC	TACAGAGATTGATTTCTTAATTAATACAGCTGAAGAACTTGAAGAAATTAACAAATTA	2082
2103	AC	AGCAACTCAAGAAAGCAGAGGTCCTTGAAGCGCGGCACTGAAGCAG	2151
2083	GC	AACTCACTCAAGAGCAAAAGTCCATGAGAGCTGAAGCACTGAAGCAG	2131

DEFINITION	Homo sapiens cDNA FLJ1940 fis. clone PLACE1010942. highly similar to Homo sapiens intersectin long isoform (ITSN) mRNA.
ACCESSION	AK027846
VERSION	AK027846.1 GI:14042823
KEYWORDS	oligo capping; fis (full insert sequence).
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	1 Isogai,T., Oca,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y., Nishikawa,T., Nagai,K., Sugano,S., Iihobashi,T., Fujimori,K., Tanai,H., Kinata,M., Watanabe,M., Hiraoka,S., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagahari,K., Masuko,Y. and Kanehorl,K.
TITLE	NEDO human cDNA sequencing project
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 2131)
AUTHORS	Isogai,T. and Otsuki,T.
TITLE	Direct Submission
JOURNAL	Submitted (10-MAY-2001) Takao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:genomices@helix.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT	NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology; cDNA library construction, 5'- & 3'- end one pass sequencing and clone selection: Helix Research Institute (supported by Japan Key Technology Center etc.) and Department of Virology, Institute of Medical Science, University of Tokyo.
FEATURES	Location/Qualifiers
Source	1..2131
	/organism="Homo sapiens"
	/mol_type="mRNA"
	/db_xref="taxon:9606"
	/clone="PLACE1010942"
	/rissue_type="Placenta"
	/clone_1fb="PLACE1"
	/note="Cloning vector: PME18SFU3"
BASE COUNT	691 a 448 c 572 g 420 t
ORIGIN	
Query Match	27.6%; Score 1404.2; DB 9; Length 2131;
Best Local Similarity	81.0%; Pred. No. 1.2e-311;
Matches 1740; Conservative	0; Mismatches 283; Indels 126; Gaps 4
CY	7 GAGAGAAGATGGAAGCGGGCCGGAGAGGGCGGCAGCTTGTGTTCTCCGTAGTACGGCGGCT 66
DB	105 GAGAAAGATGAGAGGCCAGGAGGAGGAGCCTAAGCTTGTCCTCGTAGTACGGCGGCT 164
OY	67 CGCAAGAGAGATCCCCGAGCGGGCTCCCGGAGCGGCCGGGAGGCAAGCAGCGCGCGCG 126
DB	165 CGCAGAGAAATCCCAGCGGGCTCCGGAGC-----GACAGAGAGCGCGCG 213
OY	127 GGAGTGTGTGCGCGGCTGCGACTCGGCGTTCCCTGC-GCGGCGTGCAGGCTGACCTGA 185
DB	214 GGAGTGTGTGCGGGGCTGCGGCTCTGTCCTGCTCCAGCGGGCGGTGAGCGGCACTGA 273
OY	186 TTGTGTGAGAGGGGCGCGCGCGCACCCCGCCCGGAGATGAGGCGTTCATCAGCAAGGTGA 245
DB	274 TTTGTCCCTGGGGCGGCGCGGACCAGCCGCGCGGAGATGAGGCGTTCGATTACCAAGTAA 333
OY	246 ACGTAATATGAAACAATGGCTCAGTTTCCACACCTTGGGTGTAGCTGGATGTCTGGGG 305
DB	334 AAGTACAGAACCATGCTCAGTTTCCAACACCTTTGGTGGCAGCTGGATATCTGGGG 393
OY	306 CATAACTGTGAGGAAAGGGCCAGACATGACACAGCAATGTTCTTAAGCTGAAAGCGATAAC 365
DB	394 CATAACTGTAGGAAAGGCGAAGCATGATCAGCAGTTCCATTAAGTTAAAGCAAATATC 453
OY	366 GGAATTAATTAATCTGTGATCAAGCAGAGAACTTTTTTCCATCTGGGTTACTCTGAGC 425
DB	454 TGGAATTAATTAATCTGTGATCAAGCAGAGAACTTTTTTCCATCTGGGTTACTCTGAGC 513

/db_xref="taxon:9606"
/issue_type="bone marrow"
37..1599
/product="SH3 domain-containing protein SHP1"
/protein_id="AAC50592.1"
/db_xref="GI:1438933"
/translation="MEAEERLKQKEQERKIIIELEKQEAORPAQERDKOMLEHVOQED
EHORPKLHEEKLEESYKKEGEKQEAODLGLFHOHEPAKPAVOAQMST
AEKGPLTISAQENVKVVYRALLYPESRSHDITIOGDIWMDSCOTGEPMGLGSET
KKTGWFPNPAKIPENYPAVPKVTSTSPAPKALRETPALATVSSPSTTP
NMADSTSTPSTINKEPSTMDNDAAPSLVPSAGLURQSAATPAVATCSSP
VLQGEKRVGLQDALYPMRAKQDNHINRNDVITLVEQDMAGVEYQOGKMPK
SYKLISGPIRKSTMSDSSSPASIKKVASPAKVPVSGEIIAQVIASTATGPEO
LTLAPOLILIRKQNGWMEBELQARQKROIGMPFPAVVLKLSGTSTIPTEPK
STLAACVQIGVMDYTAQNDDELAFNKQIINVLNKEPDMWKSEVNOVGLPEPNT
VLTITMD89Q"

BASE COUNT 994 a 756 c 702 g 789 t

Query Match 24.6%; Score 1252.6; DB 9; Length 3241;
Best Local Similarity 69.8%; Pred. No. 9.2e-277;
Matches 2132; Conservative 0; Mismatches 604; Indels 317; Gaps 20;

2112 CCAGAGCAGAGGTCCCTGAGGACGCGCTGAAGCAGAAAGACAGAGAGAG 2171
21 CCAGAGCAGAAAGTCCATGAGGCTGAAGCATGAACAGAAAGACAGAAAGAT 80
2172 CCTGAGTTAGAGACCAAAAGAAAGACGCTCAGAGACAGATTGAGAAAGGACAGCA 2231
81 CATGAAATTAGAAAACAAAAGAAAGAACCCAAAGACGAGCTCAGAAAGGACAGCA 140
2232 ATGGCTGAGACATGTGACAGAGAG--GAGCAGCAGCGCCCGGAAACCCACAGAGA 2288
141 GTGGCTGAGACATGTGACAGAGAGAGCAGCATCAGAGCCAAAGAACTCCACAGAGA 200
2289 GGAAGACTGAGAGAGAGAGACAGTGTCAAGAAAGAGAGAGAGAGAGAGCCAGCC 2348
201 GGAAGAACTGAAAGAGAGAGAGTGTCAAAAAGAAAGATGGAGAGAAAGGCAAAAC 260
2349 GGAAGATGCAAGACAGAGAGTGGCTTTTCCATCCGATCAGAGCCAGCTAAGTGGC 2408
261 GGAAGCAGAAAGCAAGCTGGTGGCTTTTCCATCAACCAAGAAACAGCTAAGCCAGC 320
2409 CACCCAGGACCCCTGCTTACCAAGAGAAAGCCCGCTTACATTTTTCACAGAGAG 2468
321 TGTCCAGGACCCCTGCTTACAGAGAAAGGTTCACTTACATTTTTCACAGAGAGAA 380
2469 TGTAAAGTGTATATTACCGAGCGCTGTACCCCTTTGAATCCAGAAATCAGATGAT 2528
381 TGTAAAGTGTATATTACCGGCACTGTACCCCTTTGAATCCAGAAAGCAGATGAT 440
2529 CACCATCCAGCAGAGATATAGTCATGTGTGATGAAGCCAGACTGAGAGCCAGATG 2588
441 CACTATCCAGCAGAGATATAGTCATGTGTGATGAAGCCAGAACTGAGAGACCCGCTG 500
2589 GCTTGAAGAGAGCTGAAGAGAGAGCGGATGTTCCCTGCAAAATATAGCAGAAAGAT 2648
501 GCTTGAAGAGATTAAGAGAGAGAGCGGATGTTCCCTGCAAAATATAGCAGAGAAAT 560
2649 TCCAGAAATGAGTTCCTCACTCAGCCAAACAGAGTCCAGATCTGACATTCGCCCTGC 2708
561 CCCAGAAATGAGTTCCTCACTCAGTGAACCAAGTGAATGATTCACATTCGCCCTGC 620
2709 CCCCAACTGCTCTGCTGAGAGACCCCTGCTCTTTGCAAGTACCTTTTGAAGCCCTC 2768
621 CCCCAACTGCTCTGCTGAGAGACCCCTGCTCTTTGCAAGTACCTTTTGAAGCCCTC 680
2769 CACAAACCCCAACCACTGGGAGACTTCAGTTCCAGTGGCCAGAGAGCTCAACAGAGAA 2828
681 CACGACCCCTTAATTAATGAGGAGACTTCAGTTCCAGTGGCCAGAGAGCTCAACAGAGAA 740
2829 GCCAGAAACGACCACTGGGATACGTGGGCGGCTCAGCCCTTCTGACCGTACTAGTGC 2888

741 ACCAGAAAGGATTAAGTGGATGATGAGGACAGCCCTCTCTACCGTTCCAGATGC 800
2889 TGGCCAGTTACGGCAGAGATCAAGCTTTTACCCAGCAGACGACTGCTCTCCCTC 2948
801 CGGCCAGTTAAGCAGAGATCGGCTTTTACTCCAGCAGGCGCAGCTCTCCCTCCGCT 860
2949 TCCCGCTCGGCGCAGGAGGAGAAAGGTGAGAGGAGGCTCAAGAGGCAAGCCCTGATCC 3008
861 TCTGTGCTAGGCGCAGGAGGAGAAAGGTGAGAGGAGGCTCAAGCTCAAGCTTATCTTG 920
3009 GAGAGCAAAAGAGCAACCACTTAATTTTAAAGAGTGAAGTCAAGCTTCAAGCTTGA 3068
921 GAGAGCAAAAGAGCAACCACTTAATTTTAAAGAGTGAAGTCAAGCTTCAAGCTTGA 980
3069 ACAGCAGACATGTGTGTTTGAAGAGTCAAGAGTCAAGAGGTTGTTCCCAAGTC 3128
981 ACAGCAGACATGTGTGTTTGAAGAGTCAAGAGTCAAGAGGTTGTTCCCAAGTC 1040
3129 TTACGTGAACCTCATTTCAAGGCGCGTAAAGAAATCCACAGCATGATAGGCGCTAC 3188
1041 TTACGTGAACCTCATTTCAAGGCGCGTAAAGAAATCCACAGCATGATAGGCGCTAC 1100
3189 TGAAGTCCCTGCTAGCTTAAGAGAGTGGCTTCCCGGCGGCAAGCCAGCTTCCCGG 3248
1101 AGAGAGTCCCTGCTAGCTTAAGAGAGTGGCTTCCCGGCGGCAAGCCAGCTTCCCGG 1160
3249 AGAGAGTTCATTCAGTACATACATAGAGTTCAGAGAGTTCAGAGAGATTAACCTTTCA 3308
1161 AGAA----- 1164
3309 GCAAGGAGATGATGTTGTTTACCAAGAAATGATGATCTGTGACGGAGACGGTGG 3368
1165 ----- 1164
3369 CGACAGTCCGAGTCTTCCCTTAACTATGTAGGCTTAAAGATTCAAGAGGCTGTGG 3428
1165 ----- 1164
3429 AACTGTGGGAAACAGAGGATTAAGAAACCTGAATTTGCCAGGTTATGCTTC 3488
1165 -----GAAATTGCCAGGTTATGCTTC 1187
3489 CTACGCTGCTAGTCCCGAACAATCACTGCTGCTCTGGGAGCTGATTTCTGATCCG 3548
1188 ATACACCGGACCGGCGGAGCAGCTCACTTCCCTGTGAGCTGATTTTATTCG 1247
3549 GAAAAAGAACCCAGGATGATGGTGGAGAGAGAACTGCAAGCTCGAGGAGAAAGCGCA 3608
1248 AAAAAAGAACCCAGGATGATGGTGGAGAGAGAGCTGCAAGCAGTGGAGAAAGCGCA 1307
3609 GATAGGATGTTTCCAGCAATTAATGTCAAATTTTAAAGCCCGGAAACAGCAAAATAC 3668
1308 GATAGGATGTTTCCAGCTTAATTAATGTCAAATTTTAAAGCCCTGGAAGCAAGCAATAC 1367
3669 CCCAATGATGATCCAGAGCCGAGTGAAGCAGCAGCAGCAGTGTGCTGAGTGTGATGTA 3728
1368 TCCAAACAGAGCCACTTAAGTCAACAGCATTAAGCAGTGTGCTGAGTGTGATGTA 1427
3729 CGATTACACCGGCGGAGAGATGAGAACTGAGCTTTCAGCAAAAGGCGAGATCATCAAGT 3788
1428 CGACTACACCGGCGGAGAGATGAGAGTGTGAGCTTTCAGCAAAAGGCGAGATCATCAAGT 1487
3789 CCTCAACAGAGAGAGCCGAGCTGTGAGAAAGAGAGTCAAGTGTGAGGCTCTT 3848
1488 CTTCAACAGAGAGAGCCCTGACTGTGAGAAAGAGAGTCAATGAGCAAGTGTGAGGCTCTT 1547
3849 CCCATCCAAATTAATGAAAGTGAAGCAGACAGATGAGACCCGAGCAGCAATGATATATG 3908
1548 CCCATCCAAATTAATGAAAGTGAAGCAGACAGATGAGACCCGAGCAGCAATGATATATG 1607
3909 TTGTCATTCCTCCCTCAGGCTTGAAGTCTCAAGAGAGCCCAATTCCTATCATCATG 3968

QY	4951	AGACGTGGGGGGTGTCTACAGATTAAT--ATGAATATGGCGCTCCGTCGCTGTGTGT	5008
Db	2747	GGACTGTGGGAATTTTAAATTAATTATGAAATAGAGCTCAGAGTCTTTGGGGGT	2806
QY	5009	TAAC TTGTGCTGTAGCTGAACCGCTGTCTCTTAGATTAATTAGTGAAGTCGG	5061
Db	2807	AAGAAAGCTGTAGGGGAAAGCCCTGTTGTTTAACTAGGTGAAGG	2859
RESULT 15			
LOCUS	BD127640	1676 bp	DNA linear
DEFINITION	Primer for synthesizing full-length cDNA and use thereof.		
ACCESSION	BD127640		
VERSION	BD127640.1	GI:23222585	
KEYWORDS	JP 2002017375-A/3071.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	Ota,T., Nishikawa,T., Isogai,T., Hayashi,K., Ishi,S., Kawai,Y., Makamatsu,A., Sugiyama,T., Nagai,K., Kojima,S., Otsuki,T. and Koga,H.		
TITLE	Primer for synthesizing full-length cDNA and use thereof		
JOURNAL	Patent: JP 2002017375-A 3071 22-JAN-2002;		
COMMENT	HELIX RESEARCH INSTITUTE OS Homo sapiens (human) PN JP 2002017375-A/3071 PD 22-JAN-2002 PF 07-JUL-2000 JP 2000253172 PI TOSHIO OTA,TETSUO NISHIKAWA,TAKAO ISOGAI,KOJI HAYASHI,SHIZUKO PI ISHII, PI YURI KAWAI,AI WAKAMATSU,TOMOYASU SUGIYAMA,KEIICHI NAGAI, PI SHINICHI KOJIMA, PI TETSUJI OTSUKI,HISASHI KOGA PC C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/ PC 10, C12P21/02,C12Q1/68//C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC PC Primer for synthesizing full-length cDNA and use thereof PR Key Location/Qualifiers (264).. (1676).		
FEATURES	FT	CDS	Location/Qualifiers
SOURCE		1..1676	/organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606"
BASE COUNT	485 a	381 c	458 g 352 t
ORIGIN			
Query Match	24.6%	Score 1249.2;	DB 6; Length 1676;
Best Local Similarity	85.7%;	Pred. No. 5.2e-276;	
Matches 1428;	Conservative	0; Mismatches 223;	Indels 15; Gaps 3;
QY	7	GAGGAGAGTGGAGCGGGCGGGGAGGGGCGCGAGCTTGTTGTCCTCGTAGACGGCGCT	66
Db	22	GAGAAAGATGAGGCGGCCAGGAGAGGAGACCTAGTTGTTGCTCTCGTAGACGGCGCT	81
QY	67	CGCAAGGAGATCCCGAGCGGGGCTCCGGAGCGGCCGGGAGGACGAGCGGGCGG	126
Db	82	CGCGAGGAAGATCCCGAGCGGGGCTCCGGAGCG-----GACGAGAGGCGGGCG	130
QY	127	GGATGTGTGTCCGGGCTGCGGAGCTCGGCTTCTTCGC-GGGGCTGCGGGCTCACCTGA	185
Db	131	GGATGTGTGTCCGGGCTGCGGAGCTCTGCGCTCTCCAGGCGGCGTAGACGGCACCTGA	190
QY	186	TTTGTGTAGAGGGGCGGCGCGACCGCGCCGGAGATGAGGCGCTGCATCGACAAGGTGA	245
Db	191	TTTGTCTCTGGGGGCGGCGAGCGCGAATCCGCCGAGATGAGGCGCTGCATTCGACAAGTGA	250
QY	246	ACGTATATGAACATAGCTCAGTTTCCACACTTTTCGGTGTAGCTTGATGTCTGGCG	305

Db	251	AAGTAAACGAACCAAGGCTCAAGTTTCAACACCTTTGGTGGAGGCTTGATATCTGGGC	310
OY	306	CATTAACGTGGAGAAAGGGCCAAAGCTACACAGAGTTCTTAgCTTAAGCCGATAC	365
Db	311	CATTAAGTGAAGAAAGGACAGAGATGATGACAGTTTCATAGTTTAAAGCAATATC	370
OY	366	GGGATTTTATCTAGGAGTCAAGCGGAACTTTTTTCCAAATCGGGTTCCTACGCC	425
Db	371	TGGATTCATTAAGCTGGATCAAGCTGAAACTTTTTTTTCAATCTGGGGTATCTCAACC	430
OY	426	TGCTCTAGACCAATATATGGGCGCTAGCGGACATGAATAACATGGAAGATGGATCAAGT	485
Db	431	TGTTTAAAGACAGATATGGGCACTAGCTGACATGAATATGATGGAAGATGATTAAGT	490
OY	486	GGAAATTTTCCATAGCCATGAAGCTTATCAACCTGAAGCTACAGAGATTAAGCTCCCTTC	545
Db	491	GGAGTTTTCATAGCTATGAATCAACTTATCAAACTGAAGCTACAGAGATTAAGCTCCCTTC	550
OY	546	CACACTTCCCCCTGTCTATGAAACAGCAACCAAGTGGCTATTTCAAGTGCACAGCATTTGG	605
Db	551	TGCACCTCCCCCTGTCTATGAAACAGCAACCAAGTGGCTATTTCAAGTGCACAGCATTTGG	610
OY	606	TATAGAGAGATTTGCTAGCATGCACTCAAGCTGTGTGCTCCGTGACCAATGGGCTTC	665
Db	611	TATGGAAAGTATCCGACAGCTACCAAGCTTACAGCTGTGCTCCGTCACATGGGATTC	670
OY	666	CATTCAAGTTGTTGGAAATGTCTCAACCCCTTAGATCTTCTGTCCCTCCAGCAGCAGTGGC	725
Db	671	CATTCAAGTTGTTGGAAATGTCTCAACCCCTTAGATCTTCTGTCCCAAGCAGCAGTGGC	730
OY	726	TCCCTGGCTAAAGGGGCTCTCCCTCCGTCAATACAGCTCTGCTCGGTTGGCATCTGTC	785
Db	731	CCCCCTGGCTAAAGGGGCTCTCCCTCCGTCAATACAGCTCTGCTCGGTTGGCATCTGTC	790
OY	786	AGCCACATGGCCAAAGAGTTCTTCCCTGACGAGTCCGTCAGGGTCACAATTAACAC	845
Db	791	AGCCACATGGCCAAAGAGTTCTTCCCTGATGATCTGTCAGGGTCACAATTAACAC	850
OY	846	TAACTTACAGAAAGGACCAATCATTCGATGTCCGACGCGCCCTCCAGCAGCAGAAATGGGC	905
Db	851	TAAATTAACAAAGGACCAAGTCAATTTGATGTGCGCAGTGTCCACAGTGGCAGATGGGC	910
OY	906	TGTGCTCAGTCAATCAAGGCTGAATATACAGCAGATTATTAACAGCCAGCAAAACTAT	965
Db	911	TGTTCTCAGTCAATCAAGGCTGAATATACAGCAGATTATTAACAGCCAGCAAAACTAT	970
OY	966	GAGTGAACATTAACAGGTCCCCCAGCAAGAACTATCTCATGCAATCAAGTTTACCCTA	1022
Db	971	GAGTGAACATTAACAGGTCCCCCAGCAAGAACTATCTTAAGCAGTCAAGTTTACCCTA	1033
OY	1026	GGCTCAGCTGGCTCAATATGGAATCTTTCTGACATGTATCAAGATGGAATCTACATGC	1088
Db	1031	GGCTCAGCTGGCTCAATATGGAATCTTTCTGACATGTATCAAGATGGAATCTTAACGC	1099
OY	1086	AGAAAGATTTATCTAGCTATGACACTTAATTGATGTGTCATGTCTGTCAAGCACTGTC	1144
Db	1091	AGAGAAATTTATCTGACATGCACTCATTTATGATGATATGTCTGTGACCAACTGTC	1155
OY	1146	GCCGCTCTGCTCCAGAAATATATCCCTCTCTTCAAGAAAGTTCCGTCGGCAGTGG	1207
Db	1151	ACCTGTCTGCTCCAGAAATATATCCCTCTTTTAAAGAAAGTTCCGTCGGCAGTGG	1212
OY	1206	GATGTCCTCAATTAAGCTCTTCTCTGTGGATCAGAGGCTGCTGAGGAGCCGTCGTCAGA	1262
Db	1211	TATATCTGTCAATTAAGCTCAACATCTGTAGATCAGAGGCTTACAGAGAAACCAAGTTTAA	1277
OY	1266	GGATGAGCAGCAGC--CAGAGAAAGAACTGCTGTGACATTTGAAGATTAAGAGCGGA	1322
Db	1271	AGATTAACAAACAAATTAAGAAAGAAATTAACCTGTAAAGCTTTGAAGATTAAGAGCGGA	1333
OY	1323	GAACTTCAAGCAGCAGTGTGAGCTGGAAGAGCGCCCAAGCCTTTTGAAGCAGCA	1388

Db	1331	GAACTTTGAACGTGTGCAACCTCGAATCTGGAGAAAACGAAGGCAAGCTCTCTCGAACAAGCA	1390
QY	1383	GGCGAAAAGCAGGAGCGCGGTGTGGCTTCAGGCTGTGAGCGCCCGAGTACGAGAGAGAAAAGCGC	1442
Db	1391	GGCGAAGAGAGAGAGAGCGCTCGGCGCCAGCTGTGAGACGGGCGGAGCGCAGAGAGAGAAAAGAGCG	1450
QY	1443	GGAGCGGCAGAGAGCAGAGAGCCAGCCGACGCTGTGACTGTGAGAGAGCAGCTGTGAGAAACA	1502
Db	1451	TGAGCGCCAGAGGCGAAGGCGCAAAAAGACAACCTGGAACTGTGAGAGAGCAACTGTGAAAAAGCA	1510
QY	1503	GGGGGACCTGGAGGGCGGACGAGAGAGAGAGAGAGAGAGAGAGAGAGATTCAGAGAGCGCCGAGGC	1562
Db	1511	GGGGGAGCTTACAAACGGCAGAGAGAGAGAGAGAGAGAGAGAGAGAAATTTAGAGAGCCAGAGGC	1570
QY	1563	CGCAAAAACGGGAACTGGAAAAGGCAACGCAACACTTGAATTGGGAAACGGAACCCGGAGACAGA	1622
Db	1571	TGCAAAAACGGGAACTTTGAAAAGGCAACGACAACTTGATGTGGGAAACGGAATCGAAGGCAAGA	1630
QY	1623	ACTCCTGTAATCAGAGGAACAAGGAGCGAGAGGGGCAACGTTGGTCTCTG	1668
Db	1631	ACTACTTAATCAAGAAACAAGAGCAAGAGGACATATGTTGTACTTG	1676

```
Search completed: December 4, 2003, 21:34:01
Job time : 12044.2 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 4, 2003, 15:17:39 ; Search time 23 Seconds
(without alignments)
2231.438 Million cell updates/sec

Title: US-09-674-237A-3

Perfect score: 6269

Sequence: 1 MAQFPFGSLDVVAIVE.....VGLFPSNYKLTMDPSQQ 1213

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : Issued Patents AA:*

- 1: /cgn2_6/prodata/2/1aa/5A_COMB.pep:*
- 2: /cgn2_6/prodata/2/1aa/5B_COMB.pep:*
- 3: /cgn2_6/prodata/2/1aa/6A_COMB.pep:*
- 4: /cgn2_6/prodata/2/1aa/6B_COMB.pep:*
- 5: /cgn2_6/prodata/2/1aa/PCTUS_COMB.pep:*
- 6: /cgn2_6/prodata/2/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2058.5	32.8	462	4 US-08-630-915A-38	Sequence 38, Appl
2	1388.5	22.1	509	4 US-08-630-915A-194	Sequence 194, App
3	816.5	13.0	248	4 US-08-630-915A-40	Sequence 40, Appl
4	486.5	7.8	896	1 US-08-095-737-2	Sequence 2, Appl
5	486.5	7.8	896	1 US-08-480-145-2	Sequence 2, Appl
6	486.5	7.8	896	2 US-08-477-389-2	Sequence 2, Appl
7	486	7.8	897	1 US-08-095-737-4	Sequence 4, Appl
8	486	7.8	897	1 US-08-480-145-4	Sequence 4, Appl
9	486	7.8	897	2 US-08-477-389-4	Sequence 4, Appl
10	382	6.1	1162	2 US-08-728-323A-2	Sequence 2, Appl
11	382	6.1	1162	4 US-09-298-568-2	Sequence 2, Appl
12	382	6.1	1162	4 US-09-110-399-2	Sequence 2, Appl
13	381	6.1	1898	4 US-08-056-200-94	Sequence 94, Appl
14	381	6.1	1898	2 US-08-800-644-94	Sequence 94, Appl
15	333	5.3	553	1 US-08-475-894-2	Sequence 2, Appl
16	333	5.3	553	1 US-08-484-710-2	Sequence 2, Appl
17	333	5.3	553	2 US-08-484-709-2	Sequence 2, Appl
18	333	5.3	553	3 US-08-474-697-2	Sequence 2, Appl
19	325	5.2	659	4 US-08-671-354-2	Sequence 2, Appl
20	322	5.0	59	4 US-08-630-915A-133	Sequence 133, App
21	314	5.0	62	4 US-08-630-915A-135	Sequence 135, App
22	309	4.9	56	4 US-08-630-915A-114	Sequence 134, App
23	303.5	4.8	611	4 US-09-216-393B-81	Sequence 81, Appl
24	301	4.8	1231	4 US-08-714-741-41	Sequence 41, Appl
25	298.5	4.8	864	4 US-08-714-741-40	Sequence 40, Appl
26	297	4.7	2101	1 US-08-466-390-4	Sequence 4, Appl
27	297	4.7	2101	1 US-08-470-950-4	Sequence 4, Appl

28	297	4.7	2101	1 US-08-467-781-4	Sequence 4, Appl
29	297	4.7	2101	2 US-08-483-924-4	Sequence 4, Appl
30	297	4.7	2101	3 US-09-452-294-1	Sequence 1, Appl
31	296	4.7	57	4 US-08-630-915A-136	Sequence 136, App
32	296	4.7	2101	1 US-08-195-487-4	Sequence 4, Appl
33	296	4.7	2101	5 PCT-US93-06160-4	Sequence 4, Appl
34	283	4.5	1805	1 US-07-853-913-2	Sequence 2, Appl
35	281	4.5	683	6 5210183-3	Patent No. 5210183
36	281	4.5	788	4 US-08-630-915A-30	Sequence 30, Appl
37	280.5	4.5	1375	4 US-09-722-139-2	Sequence 2, Appl
38	280.5	4.5	1375	4 US-09-721-832-2	Sequence 2, Appl
39	280.5	4.5	1375	4 US-09-721-689-2	Sequence 2, Appl
40	275	4.4	1360	3 US-09-393-569-2	Sequence 2, Appl
41	275	4.4	1360	4 US-09-579-664B-14	Sequence 14, Appl
42	275	4.4	1360	4 US-09-645-456A-34	Sequence 34, Appl
43	275	4.4	1360	4 US-09-425-324A-34	Sequence 34, Appl
44	275	4.4	1360	4 US-09-645-791-34	Sequence 34, Appl
45	274.5	4.4	1354	3 US-08-685-871-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-08-630-915A-38
; Sequence 38, Application US/08630915A
; Patent No. 6309820
; GENERAL INFORMATION:
; APPLICANT: SPARKS, Andrew B.
; APPLICANT: HOFFMAN, No. 6309820h
; APPLICANT: KAY, Brian K.
; APPLICANT: FOWLES, Dana M.
; APPLICANT: MCCONNELL, Stephen J.
; TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
; TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
; NUMBER OF SEQUENCES: 227
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/630,915A
; FILING DATE: 03-APR-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 1101-174
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNTE
; INFORMATION FOR SEQ ID NO: 38:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 462 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: Peptide
; US-08-630-915A-38
Query Match 32.8%; Score 2058.5; DB 4; Length 462;
Best Local Similarity 79.3%; Pred. No. 4.8e-138;
Matches 391; Conservative 13; Mismatches 16; Indels 71; Gaps 1;

```

QY 721 PMSTTEKPLTISAQSVKVVYVYRALYPFESRSHDEITIQGDIYVWDESGTGERGWI 780
DB 41 FGLPQKKPLTISAQENKVVYVYRALYPFESRSHDEITIQGDIYVWDESGTGERGWI 100
QY 781 ELKGTGFPANVAKIPENEVEPTPAKPVDTLTSAPAKLALRETPALPTTSSEPTTP 840
DB 101 ELKGTGFPANVAKIPENEVEPTPAKPVDTLTSAPAKLALRETPALPTTSSEPTTP 160
QY 841 NNMADFSSTWPSSEKPEETDNDMTAAOPSLTVPASQOLRORSAFTPATATGSSPSVL 900
DB 161 NNMADFSSTWPSSEKPEETDNDMTAAOPSLTVPASQOLRORSAFTPATATGSSPSVL 220
QY 901 GQGEVEBELQALYPMRAXKONHLENKSDVITVLBOODMMWGEVQOGMPKSYVX 960
DB 221 GQGEVEBELQALYPMRAXKONHLENKSDVITVLBOODMMWGEVQOGMPKSYVX 280
QY 961 LISGVRKSTSIDTPTESPASLKRVASPAKPAIPGEEFLAMYESSBOGDLTFQOQD 1020
DB 281 LISGVRKSTSIDTPTESPASLKRVASPAKPAIPGEEFLAMYESSBOGDLTFQOQD 318
QY 1021 VIVVTKKQDMWMTGVGKSGVFPNSVYRLKDSGSGTAGKTSLGKKPELIAQVIASVAA 1080
DB 319 -----ELIAQVIASVAA 329
QY 1081 TGPBOLTLAPQOLILIRKKNPGWMEGELQAGKKROIGMPANVYKLSGTSKITPTE 1140
DB 330 TGPBOLTLAPQOLILIRKKNPGWMEGELQAGKKROIGMPANVYKLSGTSKITPTE 389
QY 1141 LPTAVQAVCVIGIMDYTAQNDELAFSKQIINVINKEDPDWKGVEVQVGLFPSN 1200
DB 390 LPTAVQAVCVIGIMDYTAQNDELAFSKQIINVINKEDPDWKGVEVQVGLFPSN 449
QY 1201 YVKLTMDPSQO 1213
DB 450 YVKLTMDPSQO 462

```

RESULT 2

```

US-08-630-915A-194
; Sequence 194, Application US/08630915A
; Patent No. 6309820
; GENERAL INFORMATION:
; APPLICANT: SPARKS, Andrew B.
; APPLICANT: HOFFMAN, No. 6309820h
; APPLICANT: KAY, Brian K.
; APPLICANT: FOWLES, Dana M.
; APPLICANT: MCCONNELL, Stephen J.
; TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
; TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
; TITLE OF INVENTION: USING SAME
; NUMBER OF SEQUENCES: 227
; CORRESPONDENCE ADDRESS:
; ADDRESS: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/630,915A
; FILING DATE: 03-APR-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Mirock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 1101-174
; TELECOMMUNICATION INFORMATION:

```

```

; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 194:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 509 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; US-08-630-915A-194

```

```

Query Match 22.1%; Score 1388.5; DB 4; Length 509;
Best Local Similarity 48.9%; Pred. No. 1.9e-90;
Matches 287; Conservative 85; Mismatches 124; Indels 91; Gaps 14;

```

```

QY 635 ERKSELEKQK---EDARRVQERDKWLEHVOEOPRRKHEEDRLKREDSVRKE 690
DB 6 ERKSELEKQKLEDEARAKAKQKQENIMKENLKEKE-----EKQRLQEKTEQKIQ 59
QY 691 AERAKPEMDQKSRLEPHQEPKALQAPWSTTEKPLTISAQSVKVVYVYRALYPF 750
DB 60 AERAKPEMDQKSRLEPHQEPKALQAPWSTTEKPLTISAQSVKVVYVYRALYPF 86
QY 751 SRSHDEITIQGDIYVWDESGTGERGWIAGELKGTGFPANVAKIPE--ENEVPTPAK 808
DB 87 ARNDEWSPNSGDIITQDEKTVGEPGMLYSGFQGNFPMFCNYEKPPSSENE----- 139
QY 809 VTDLTSAPAPKALRETPALPTTSSEPTTPNNMADFSSTWPSSEKPEETDNDMTAA 868
DB 140 -----KAVSPKAL-----LPTVLSLATS-----TSSEPLSNQOPASVTDYON-VS 180
QY 869 QPSLTVPASQOLRORSAFTPATATGSSPSVPLTGQGEVEBELQALYPMRAXKONHLEN 928
DB 181 FSNLTVNTSWQ--KSAFRTVSPG--SVSPHGGQVVENLKQALCSWAKKUNHLENFS 237
QY 929 KSDVITVLEOODMMWGEVQOGMPKSYVYKLSGVRKSTSIDTPTESPASLKRVAS 988
DB 238 KHDITVLEOODMMWGEVQOGMPKSYVYKLSGVRKSTSIDTPTESPASLKRVAS 289
QY 989 --PAKPAIPGEEFLAMYESSBOGDLTFQOQDVIVTKKQDMWMTGVGKSGVFPNS 1046
DB 290 KKPISAAYSVGEVYLYPYSSVEPGDLTFEGEELIVTQDGBWMTGSGDRGIFPSN 349
QY 1047 YVRLKDSGSGTAGKTSLGKKPELIAQVIASVYATGSEQLTLAPQOLILIRKKNPGWME 1106
DB 350 YVRLKDSGSGTAGKTSLGKKPELIAQVIASVYATGSEQLTLAPQOLILIRKKNPGWME 409
QY 1107 GELQAGKQIIGFMPANVYKLSPTGSKITPTELPTAVQAVCVIGIMDYTAQNDEL 1166
DB 410 GELQAGKQIIGFMPANVYKLSPTGSKITPTELPTAVQAVCVIGIMDYTAQNDEL 462
QY 1167 LAFSKQIINVINKEDPDWKGVEVQVGLFPSNYVYKLTMDPSQO 1213
DB 463 LAFSKQIINVINKEDPDWKGVEVQVGLFPSNYVYKLTMDPSQO 509

```

RESULT 3

```

US-08-630-915A-40
; Sequence 40, Application US/08630915A
; Patent No. 6309820
; GENERAL INFORMATION:
; APPLICANT: SPARKS, Andrew B.
; APPLICANT: HOFFMAN, No. 6309820h
; APPLICANT: KAY, Brian K.
; APPLICANT: FOWLES, Dana M.
; APPLICANT: MCCONNELL, Stephen J.
; TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
; TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
; TITLE OF INVENTION: USING SAME
; NUMBER OF SEQUENCES: 227
; CORRESPONDENCE ADDRESS:
; ADDRESS: Pennie & Edmonds LLP

```

```

STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/630,915A
FILING DATE: 03-APR-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Mirock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-174
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 40:
SEQUENCE CHARACTERISTICS:
LENGTH: 248 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-630-915A-40

```

```

Query Match          13.0%  Score 816.5; DB 4; Length 248;
Best Local Similarity 58.4%; Pred. No. 2.6e-50;
Matches 153; Conservative 38; Mismatches 54; Indels 17; Gaps 3;

QY 954 PFKSVYKLGSPVKRSTSIDTPTSPASLKRVAS--PAKPAIDPEEFIAMYTESSSQ 1011
DB 2 FFAASTKILPGSEVK-----REPEALYAANVKKPSAAYSGEETIATLPPSSVVP 53

QY 1012 GDLTQOGDVIYVTKKDGDMWTGVDKSGVFPSPYVRLKDSGSGTAGKTSGLKKPEI 1071
DB 54 GDLTFTGEELIVTKDGEWMTGSIIDRSIGFPSPYVKKPKDQESFGSASKSGASNNKPEI 113

QY 1072 AQVIVSYATGPEOGLTARGLILIRKKRPGGMMWEGELQARGKKQIGFPPANVYKLLSP 1131
DB 114 AQVITAYVASGSEQLSLAPGLILIKKNTSGMWGELQARGKKQGFPPASHVYKLLGP 173

QY 1132 GTSKITPELPTAVQPAVCQVIGMYDYTAQNDDELAFSGQIINVLNKEDPDWKGVEVS 1191
DB 174 SERRATPAHP-----VCQVIAMVDYANNDELSPFGQLINVMNKDPPDMQGEIN 226

QY 1192 GQVGLFPSPNYVYKLLTMDPSQQ 1213
DB 227 GVTGLFPSPNYVMTTDSPSQQ 248

```

```

RESULT 4
US-08-095-737-2
; Sequence 2, Application US/08095737
; Patent No. 5487979
; GENERAL INFORMATION:
; APPLICANT: DiFiore, Pier P
; TITLE OF INVENTION: A Substrate for the Epidermal Growth
; TITLE OF INVENTION: Factor Receptor Kinase
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 620 Newport Center Drive, Sixteenth Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: United States of America
; ZIP: 92660

```

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/095,737
FILING DATE: 19930722
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Israelson, Ned A
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH060.001A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 896 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-095-737-2

```

```

Query Match          7.8%  Score 486.5; DB 1; Length 896;
Best Local Similarity 20.5%; Pred. No. 4.3e-26;
Matches 198; Conservative 150; Mismatches 273; Indels 345; Gaps 29;

QY 15 WAITVERAKHDQDELSTKPIAGFITDQARNFFQSGLPQVLAQIWMALADNMNDGRMD 74
DB 122 WAKVEDRAKVDALPDSLSPVNGFLSGDKVKRVKVLNLSKLPVILGRVWELSDIDDMGD 181

QY 75 QVESITAMKLYLKLQGIQSLSTLPVWKQDPVAISSAPAFSIGIASHMPILTAAPVPM 134
DB 182 RDEFVAMFLVYCALE-----XEPVPM 203

QY 135 GSIPVGMSPPLVSSVPPAAPPPLANGAPVPIQPLPAFAHPAATPKSSFSRSGPQSL 194
DB 204 -----SLPPLVPP-----SKR 215

QY 195 NTKLQKASFDVASAPPAEAMVAPPOSSRLKYROLFNSHDKTWSGHLTGPOARTIMOSSL 254
DB 216 KT-----WVSPAEKAKYDEILFKTDKMDGVSGLVGEVLEIFLKTGL 257

QY 255 POAQLASTWNISDIDQDKLTAEFTLMLHLDVAM-SGQPLPVLPPRYIPPSRRVRS 313
DB 258 PSTLAAHTWSLCDTDCCKLSKQCFALAFHLISQKLIGIDPPHYLTPEMTIPPS--DRA 314

QY 314 GSGMVISSSSYDQRLPEPSEDEQPEKKLPVTEEDKKRNFERSGVELEKRRQALIE 373
DB 315 SLQKNIISSPV-----ADPSAIK-----ELDTLNNEIVD 344

QY 374 QQRKEQRLAQLERAQERKEREROEQAQTLELEKQRELEPQREERREIERR 433
DB 345 LQREK-----NNVEQDLKEKE-----DTIKQRTSEVQ--DLQDEVQRE 380

QY 434 EAKKELEQRLQLENNRRQELLNQRNKEQEGTVLAKRRKTLFELBALNDKKHQLQEG 493
DB 381 NTNLOKLOAQKQ-----QVQELIDE-----LDEQAQLELQ 410

QY 494 KLQDIRCLATROEISTNKSRLRIETHTLQOOLQESQOMGLRILPEKILSDQKQ 553
DB 411 QKEVKKCAEBAQILSLK-----ABLTQESQISTYEELAK-----ADELSR 456

QY 554 VQONSILHRDLSLTTLKRALEAKELAQQLREQLDEVERETR-----KLQEDIVFNNOQL 607
DB 457 LQQETAE-----LEESVESGAQLEPLQGHQDSQQLSSQMKMLMKDLNNNSQLN 510

QY 608 ELREIHS-----KQOLQRELEAARLQKQERKSLTL----- 641
DB 511 WCSSPHSLIVNGATDYCSLSTSSSETANLNEHVQESQSLSESPRIQESPPASSPILPSG 570
QY 642 EKQKEDAQRVQERKQVLEHVQEEQRPKPKHEDELDKEDSVRKKEABER----- 694

```

```

Db      571 VTDENEVTTATYKVCSELDN-----NRHSKEEDPFNVDSSTLNGPVDNLDFPQS 622
Qy      695 -----AKPEMODKOSRLFHPHOEPK-----LATQAPWSTTEKGLTISAQESV 738
Db      623 DPFVGSDFPKDDPFKIDPFQGDPPKSDPPASDCFFRQSTDPFATSTDPFSAANSSI 682
Qy      739 KVVYRRALYPFESHDEITIQPGDIYVWDESQT-----GEQWJCG-----780
Db      683 TSV-----ETLKNDPFAFGTVAASDASADDPFASVFGNSFGGFPADFSTLSKVN 734
Qy      781 ---ELKGTGMFPAN-----YAEKIPENEVPTPAKPVTDLTSAAPKALRETAPLP 830
Db      735 NEDPRSATSSSVSNVITKAVFETSVKSEDEPPALP-----PKIGTPTRPCPLP 785
Qy      831 -----VTSSEP-----STTPNNWADPSSITWSSSNEKPETDWMT 865
Db      786 PGKRSINKLDSPPDKLNDPFQPPGNDSPKXKPEMFCDPFTSATTTNKADPSPNFAN 845
Qy      866 WAAOPS 871
Db      846 FSAIYS 851

RESULT 5
US-08-480-145-2
; Sequence 2, Application US/08480145
; Patent No. 5717067
; GENERAL INFORMATION:
; APPLICANT: DiFiore, Pier P
; APPLICANT: Fazioli, Francesca
; TITLE OF INVENTION: A Substrate for the Epidermal Growth
; TITLE OF INVENTION: Factor Receptor Kinase
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 620 Newport Center Drive, Sixteenth Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: United States of America
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,145
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/095,737
; FILING DATE: 22-JUL-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelien, Ned A
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH060.001A
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 896 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-480-145-2

```

```

Query Match      7.8%; Score 486.5; DB 1; Length 896;
Best Local Similarity 20.5%; Pred. No. 4.3e-26;
Matches 198; Conservative 150; Mismatches 273; Indels 345; Gaps 29;
Qy      15 WAITEEBAKHDOQLSLKPIAGFITGDQARVFFQSGLPQVLAQIVMALDMNNDGRMD 74

```

```

Db      122 WAVEPEDKAKYDAIFDLSLVNGLSGDYKVPYLANSKLVDLIGRWELSDIDHGMLD 181
Qy      75 QVEESIMAKLIKLLQGYQLPSTLPVWKQOPVAISSAPFGIGLISMPELTVAAYPM 134
Db      182 RDEFAVAMFLVYCALE-----KEEVPW 203
Qy      135 GSIVVGMSPPLVSVVPAAPVPLANGAPVVIQPLPAFAHAPATWPKSSPSRSGPSOL 194
Db      204 -----SLPPALVPP-----SKR 215
Qy      195 NTKLQKAQSPDVASAPPAEMAVPQSSRLKYRQLFNSHDTWSGHLTGPOARTIMQSL 254
Db      216 KT-----WVVSAPAKAYDEFLTDMQMGVFGLEVRBELFLKTGL 257
Qy      255 PQQLASITWNLSDIDQCKLTAEFFILAMHLIDVAM--SGQPLPVLPPEVTPPEFRVRS 313
Db      258 PSTLAAHWISLCTDKDCKLSKQFALAFHLSQKLKIGIDPPHVLTPPEMIPPS---DRA 314
Qy      314 GSGMSVSSSSVDQRLPEEPSSEDEQPEKKLPVTFEDKKRENERGSEVELEKROALLE 373
Db      315 SLQKNITIGSEPV-----ADFSAIK-----ELDTIANEIVD 344
Qy      374 QQRKEERLAQLERAQERKERERQEQAKROLELEKQLEKORELERQREERKEIERR 433
Db      345 LQREK-----NVVEDLKEKE-----DTIKORTSEVQ---DLQDEVQRE 380
Qy      434 EAARKELEERQOLEMENRROELNQRNKEQEGTVLKARKRTLEFLEALNDKQHOLEG 493
Db      381 NTNLQKLOAQKO-----QVQELDE-----LDEQKALEB 410
Qy      494 KLDIRCRLATQROEISTKSRRLAETIHLQOQLQESQWGLRILPEKQILSDQK 553
Db      411 QLKEVRKKCAEAOIISLK-----AELTQESQSTYEBELAK-----ARELSR 456
Qy      554 VQONSILRDSILTLKRALLEAKELARQQLREQLDEVERETR-----KLQETIDVFNQK 607
Db      457 LQGETAE-----LEESVSGKQLERPLQHLQDSQOEISMQWKLMEKMLEHNSQLN 510
Qy      608 ELREIHS-----KOQLOKQSLAARLKOQEOKSLEL-----641
Db      511 WCSSPHSILVNGATDYCSLTSSETANLNEHVGQSNLESEPIHOSPPASSPELLPSG 570
Qy      642 EKQKEDAQRYQEBDKQMLEHVQOEBQPRPKPHEERLKRKEDSVRKKAEER-----694
Db      571 VTDENEVTTATYKVCSELDN-----NRHSKEEDPFNVDSSTLNGPVDNLDFPQS 622
Qy      695 -----AKPEMODKOSRLFHPHOEPK-----LATQAPWSTTEKGLTISAQESV 738
Db      623 DPFVGSDFPKDDPFKIDPFQGDPPKSDPPASDCFFRQSTDPFATSTDPFSAANSSI 682
Qy      739 KVVYRRALYPFESHDEITIQPGDIYVWDESQT-----GEQWJCG-----780
Db      683 TSV-----ETLKNDPFAFGTVAASDASADDPFASVFGNSFGGFPADFSTLSKVN 734
Qy      781 ---ELKGTGMFPAN-----YAEKIPENEVPTPAKPVTDLTSAAPKALRETAPLP 830
Db      735 NEDPRSATSSSVSNVITKAVFETSVKSEDEPPALP-----PKIGTPTRPCPLP 785
Qy      831 -----VTSSEP-----STTPNNWADPSSITWSSSNEKPETDWMT 865
Db      786 PGKRSINKLDSPPDKLNDPFQPPGNDSPKXKPEMFCDPFTSATTTNKADPSPNFAN 845
Qy      866 WAAOPS 871
Db      846 FSAIYS 851

```

```

RESULT 6
US-08-477-389-2
; Sequence 2, Application US/08477389
; Patent No. 5872219
; GENERAL INFORMATION:

```

```

APPLICANT: DiFiore, Pier P
APPLICANT: Fazioli, Francesca
TITLE OF INVENTION: A Substrate for the Epidermal Growth
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESS: Knobbe, Martens, Olson & Bear
STREET: 620 Newport Center Drive, Sixteenth Floor
CITY: Newport Beach
STATE: California
COUNTRY: United States of America
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/477,389
FILING DATE: 07-JUN-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/095,737
FILING DATE: 22-JUL-1993
ATTORNEY/AGENT INFORMATION:
NAME: Israelien, Ned A
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH060,001A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 896 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-477-389-2

```

```

Query Match 7.8%; Score 486.5; DB 2; Length 896;
Best Local Similarity 20.5%; Pred. No. 4.3e-26;
Matches 198; Conservative 150; Mismatches 273; Indels 345; Gaps 29;

```

```

15 WAITVEERAKHDOQFLSTKPIAGFLTGDDARBNFPQSGLPQVLAQIMALAMNNGRD 74
122 WAVEKEDRKAKYDAIFPSLSPVNGFLSGDVKPVLSKLPVLDLGRWELSDIDHGMID 181
75 QVEFSIAMLKIKLQGYQLPSTLPVWMOQPAVIASSAPAFGIGIASMPILAVAPVM 134
182 RDEPAVAMFLVYCALE-----KEPVPM 203
135 GSIPVVGMSPLVSVPPAAVPPPLANGAPVVIQPLPAFAHPAATWPKSSFSRSGSQ 194
204 -----SLPPLVPP-----SKR 215
195 NTKLQAKGSFDVASAPPAEMVPOSSRLKYROLFNSHDKTWSGHLTGPQARTIMQSSL 254
216 KT-----WVSPAKKAKTDEIFLTKDKMDGFGVSGLEVEIFLTKL 257
255 PQAOLASINWLSIDIDODGKLTAEFPIAMHLIDVAM-SCQPLPVVLPPEYIPPSFRVRS 313
258 PFTLLAHIMSLCTDKOCGLSKDQFALAFHLISQKLIKIDPPHVLTPMIPPS---DRA 314
314 GSGMSVSISSVDQRLPEPSSSEDEQPEKKLPVTFEDKKRENFERSGVELEKRRQALIE 373
315 SLQKNIIGSSPV-----ADPSAIK-----ELDTLNNEIYD 344
374 QQRKEQERLAQLEBAQERKERERQEAQOLELEKQLEKOREEERERKXIERR 433
345 LQREK-----NNVQDLKEKE-----DTTKQTSSEVQ---DLQDEVQRE 380
434 EAKKELEERQOLEWERNRROELANDNRNKEOGTVVLKARKRTLEFELBALNDKKHQLRG 493

```

```

DB 381 NTKLQKQAQKQ-----QVCELDE-----LDEQKAQLE 410
494 KQDRCRAQROBIESTNKRELRIAEITHLQOOLQESQOMGLRIPKQILSQKQ 553
411 QKEVRKKCAEAQILSLK-----ABLSQESQISTYEBELAK-----ABEELR 456
554 VQONSLHRDILLTLRALAEAKELAAQOLREQUDEVERETR-----KLQEDIVFNQK 607
457 LQQETAE-----LEASVESGKAQLEPLQCHQDSQQLSSQMOMLWEMKDLNNNSQL 510
608 ELREIHS-----KQOLQKRSLEAARLKQKEQERSLEL--- 641
511 WCSSPHSILVNGATDYCSLSTSSSETAMINEHEVQCSNLESEPIHQESPARSSPELPSG 570
642 EKQKDQARRQERQKQMLEHQEQEQRPRKPHEDRLKEDSVRKKAEER----- 694
571 VTDENEVTTAVTEKCSIEDN-----NRHSKEEDPENVSSILTGFAVATNLDPFOS 622
695 -----AKPEMOKQGRLEFHPQEPAP-----LATQAPWSTTEKPTLISQESV 738
623 DPFVSDPFKDDPFKIDPFQDPPKSGDPFASDCFRQSTDPFATSSDTPFSANNSI 682
739 KVVYRALYPFESRSHDEITTPGDIWVDESQT-----GEPQWLQ----- 780
683 TSV-----ETLKHNDPFAPGTVVAASDSDATDPFASVFNESFGGFADPSTLSKVN 734
781 -----ELKGTGWFPAN-----YAKIPENEVPTPAKPTDLSAPAPLARETPAP 830
735 NEDPFRSATSSSVSNVITKNVFEETSVKSEDEPPALP-----PIGTPTRCPLP 785
831 -----VTSEEP-----STPNNMADFSSTWSSSENEKPEETNMOT 865
786 PGRSINKLSDSPDKLNDPFPFPFGNDSPKKXDEPMFCDDPTSATTTINKADPSNFPAN 845
866 WAAQPS 871
846 FSAVPS 851

```

```

RESULT 7
US-08-095-737-4
Sequence 4, Application US/08095737
Patent No. 5487979
GENERAL INFORMATION:
APPLICANT: DiFiore, Pier P
APPLICANT: Fazioli, Francesca
TITLE OF INVENTION: A Substrate for the Epidermal Growth
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESS: Knobbe, Martens, Olson & Bear
STREET: 620 Newport Center Drive, Sixteenth Floor
CITY: Newport Beach
STATE: California
COUNTRY: United States of America
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/095,737
FILING DATE: 19930722
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Israelien, Ned A
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH060,001A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 4:

```


SEQUENCE CHARACTERISTICS:
 LENGTH: 897 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-095-737-4

Query Match 7.8%; Score 486; DB 1; Length 897;
 Best Local Similarity 21.3%; Pred. No. 4.7e-26;
 Matches 205; Conservative 138; Mismatches 298; Indels 322; Gaps 31;

```

QY 15 WAITVERAKHDOQFLSLKPIAGITGDOANFFQSGLPQVLAQIWAADMMNDGMD 74
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 122 WAVESEDAKAYDAFDLSFVDGFLSGDKVAVLNSKLPEILIGRWELSDIDHDKLD 181
QY 75 QVESIMAMKLIKILKQGYQLPSTLPVWKQOPVAISSAPAGIGIASMPPLTAVAPVM 134
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 182 RDEFAVAMFLVYCALE-----KEPYPM 203
QY 135 GSIPVGMSPPLVSSVPPAAVPLANGAPVIGLPAPFAHPAATWPKSSFSRSGPSQL 194
    ||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 204 -----SLPPALVPP-----SKR 215
QY 195 NTKLQKQSFVVASAPPAEMAVPQSSRLKYROLFNSHDKTMSGHLTGPOARTILMOSSL 254
    ||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 216 KT-----WVSPAERAKAYDEIFLKTDKMDGVSGLEVRFTFLTKGL 257
QY 255 POAOLASIMLSDIDODGKLTAEFFILAMHLIDVAM-SGQPLPVLPEYIPSPFRVRS 313
    ||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 258 PSALLAHIMSLDITGCGKLSKQDFALAFHLINQKIDPPHSLTPKEMIPPS---DRS 314
QY 314 GSGMSVSSSSVDQRLPEPSSSEDEQPEKKLPVTEDEKKEKNEPERSVELEKXKQALAE 373
    ||:|||||:|||||:|||||:|||||:|||||:
Db 315 SLQKNITGSSPV-----ADPSAIK-----ELDTLNNEITVD 344
QY 374 QQRKEOERLAOLEAEOERKEREROEAKQLELEKOLEKQERELEROEERKEIERR 433
    ||:|||||:|||||:|||||:|||||:|||||:
Db 345 LQREK-----NNEQDLKEKE-----DTVQQRSEVQ---DLQDEVQRE 380
QY 434 EAARKELEERQOLEWERNRROELLNQRNKEQEGVVLKARKKLTLEFELALNDKKHOLEG 493
    ||:|||||:|||||:|||||:|||||:|||||:
Db 381 SINQKIQAKQKQ-----QVQELLGE-----LDEQKALQEE 410
QY 494 KLQDIRCRLATQROEISTNSRRLTAETHLQOQLQESQOMGLRLEKQILSDQKQ 553
    ||:|||||:|||||:|||||:|||||:|||||:
Db 411 QLOEVRKKCAEAQIISLK-----AEITQSQSISSYEELIK-----ARELSR 456
QY 554 VQONSILHSDSLITLKRALEAKELARQQLREQLDEVERE-----TRSKLOEIDVENNOIKE 608
    ||:|||||:|||||:|||||:|||||:|||||:
Db 457 LQGETAQ-----LEESVESGKAQLEPLQCHLQESQOQIISSMQKRLKMDLETNNQ--- 507
QY 609 LREIHSKQQLQKQ-----RSLEPAARLKQKEQERKSLLEKQKEDAQRRVQ 653
    ||:|||||:|||||:|||||:|||||:|||||:
Db 508 -----SNWSSSPQSVLVNGATDYCSLTSSETANFNHAGQNNLESEPHQSSSVRS 562
QY 654 ER-----DKQWLEHVQGEQOPRRPKHHEEDRLKREDSVAKKA-----ERRAPENQDK 702
    ||:|||||:|||||:|||||:|||||:|||||:
Db 563 PELAPSDVTDESEAVTAVAGNEKVTPR--FDDDKISKEDPPNVSSSITDVAATNIDDF 620
QY 703 QSRLF-----HPHOEPAKLATQAWSTTEKGLTISQESQKVVY 743
    ||:|||||:|||||:|||||:|||||:|||||:
Db 621 QSDDFVSGDPPKIDPFKIDPFQGDPRFKGSDPFASDCFFKQTSIDPFTTSSDTPFSASN 680
QY 744 RALYPFESRSHDEITIQGDIVMWDESQTSR--CWLGELKKGITGMPFANYAEKIPENE 801
    ||:|||||:|||||:|||||:|||||:|||||:
Db 681 SSNTSVETWKMNDPAPGTVVAAAASDATDPFASVFQNESFG--DGFADFTLSKVNND 739
QY 802 VPTP-----AKVTDLTSA-----PA--PKALAEETAPLPVTSSESTPPNN 842
    ||:|||||:|||||:|||||:|||||:|||||:
Db 740 AFNFTLSSSTSSVTIAKPMLEETASKSDVPALPAPKGTPTRPCPP-----FGKRRIN 794
QY 843 WADESSST-----WPSSSNEKPEITDNDMTWAAQPSLTVPSAGQLRQSAFTPATATGSS 895
    ||:|||||:|||||:|||||:|||||:|||||:

```

Db 795 KLDSDDLKLNDRPQPPGNDSPKEXQPDMEFC-----DPFTSSTTNKE 838
 QY 896 PSP 898
 Db 839 ADP 841

RESULT 8
 US-08-480-145-4
 ; Sequence 4, Application US/08480145
 ; Patent No. 5717067
 ; GENERAL INFORMATION:
 ; APPLICANT: Diflore, Pier P
 ; APPLICANT: Razoli, Francesca
 ; TITLE OF INVENTION: A Substrate for the Epidermal Growth
 ; FACTOR RECEPTOR KINASE
 ; NUMBER OF SEQUENCES: 4
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Knobbe, Martens, Olson & Bear
 ; STREET: 620 Newport Center Drive, Sixteenth Floor
 ; CITY: Newport Beach
 ; STATE: California
 ; COUNTRY: United States of America
 ; ZIP: 92660
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/480,145
 ; FILING DATE: 07-JUN-1995
 ; CLASSIFICATION: 530
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/095,737
 ; FILING DATE: 22-JUL-1993
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Israelsen, Ned A
 ; REGISTRATION NUMBER: 29,655
 ; REFERENCE/DOCKET NUMBER: NIH060.001A
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (619) 235-0176
 ; TELEFAX: (619) 235-0176
 ; INFORMATION FOR SEQ ID NO: 4:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 897 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-480-145-4

Query Match 7.8%; Score 486; DB 1; Length 897;
 Best Local Similarity 21.3%; Pred. No. 4.7e-26;
 Matches 205; Conservative 138; Mismatches 298; Indels 322; Gaps 31;

```

QY 15 WAITVERAKHDOQFLSLKPIAGITGDOANFFQSGLPQVLAQIWAADMMNDGMD 74
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 122 WAVESEDAKAYDAFDLSFVDGFLSGDKVAVLNSKLPEILIGRWELSDIDHDKLD 181
QY 75 QVESIMAMKLIKILKQGYQLPSTLPVWKQOPVAISSAPAGIGIASMPPLTAVAPVM 134
    ||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 182 RDEFAVAMFLVYCALE-----KEPYPM 203
QY 135 GSIPVGMSPPLVSSVPPAAVPLANGAPVIGLPAPFAHPAATWPKSSFSRSGPSQL 194
    ||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 204 -----SLPPALVPP-----SKR 215
QY 195 NTKLQKQSFVVASAPPAEMAVPQSSRLKYROLFNSHDKTMSGHLTGPOARTILMOSSL 254
    ||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 216 KT-----WVSPAERAKAYDEIFLKTDKMDGVSGLEVRFTFLTKGL 257
QY 255 POAOLASIMLSDIDODGKLTAEFFILAMHLIDVAM-SGQPLPVLPEYIPSPFRVRS 313
    ||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

Db 258 PSALLAHIMSLCDTKGCKGLSKDQFALAFHILINQKIGIDPESHITPEMIPPS---DRS 314
Qy 314 GSGMSVYSSSVQDRLPEERSSSEDEQPEKLPVTEDEKRENFERSVELEKROALLE 373
Db 315 SLQKNTGSSPV-----ADSSAIK-----ELDTLNEIYD 344
Qy 374 QORKEQERLAQLERAEQRERERQOEAKROLELEKOLEKOREROREEERKEIER 433
Db 345 LOREK-----NNVQDLKEKE-----DTVKORTSEVQ---DLQDEVORE 380
Qy 434 EAKRELEROROLEWERNROROLLNQRNKEQEGTVLAKARKLTLEFELALNDKQOLEG 493
Db 381 SINLOQLQAKOQ-----OVQELIGE-----LDQKQALE 410
Qy 494 KLQDRCRATQROEIESTNKSRELIAETIHLQOOLQESQOMLGLRLEPEKOLISQOLQ 553
Db 411 QLOEVAKKAEBAQOLISSLK-----AETSOESQISSYEBELK-----AREELR 456
Qy 554 VOONSILHRDLSLTKRALFAKELARQOLREOLDEVERE-----TRSKLOEIDVFNQLEK 608
Db 457 LOQETQO-----LEESVESGKQOLEPLQOHLQESQOEISSMQMRLEMDLETNNQ--- 507
Qy 609 LREIHSKQOLQO-----RSLFAARLKQOEQRKSLLEKQEKEDAQRVQ 653
Db 508 -----SNMSSSPQSVLVNGATDYCSLSTSSSETANFNEHAGQNNLESEPTHOESSVRSS 562
Qy 654 ER-----DKQMLEHVQOEQPRPKRPHEDRLKREDSVRKKEA-----EERAKPEMQDK 702
Db 563 PEIAPSDVTDEBAVTVAGNEKVTPR--FDDDKHSKEEDPFNVESSSLTDAVADTNLDPF 620
Qy 703 QSRLF-----HPOEPAKLATQAPWSTTEKGLTISAQSVAVVY 743
Db 621 QSDPFQSGDPFKDDPFKGDGDPFKGSDPFASCFKQISTDPTTSSIDPFASASN 680
Qy 744 RALYPESHSDEITIQGDIVWVDSQTEGP--GMLGGLKQKTMFPANVAEKIPENE 801
Db 681 SNTSVETWKHNDPFAFGTVVAASDASATDPFASVFGNESFC--DGFADPSTLSKYNNED 739
Qy 802 VETP-----AKPTDLTSA-----PA--PKLAIRETPALPVTSSSEPTTPNN 842
Db 740 AENPTISSSTSVTLKPMLEETASKSEBVPALPKVGPTRPCPP-----PGKRPAIN 794
Qy 843 MADFEST-----WPSSNEKPEETDNPWTMAQPSLTVPSAQQLAQRSAFPATATGSS 895
Db 795 KLDSSDPLKLNDFPQDFPQNDSPKXKOPDMFC-----DPTSSITTNKE 838
Qy 896 PSP 898
Db 839 ADP 841

```

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/477,389
FILING DATE: 07-JUN-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/095,737
FILING DATE: 22-JUL-1993
ATTORNEY/AGENT INFORMATION:
NAME: Israelien, Ned A
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH060.001A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 897 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-477-389-4

Query Match 7.8%; Score 486; DB 2; Length 897;
Best Local Similarity 21.3%; Pred. No. 4.7e-26;
Matches 205; Conservative 138; Mismatches 298; Indels 322; Gaps 31;

Qy 15 WAIYTERAKHQOFLSKPIAGFTTGQARNFPOSGILPQVLAQIWMADMMNDGRND 74
Db 122 WAKSEDKAKYPAIDPSLSPVGDGKVKVLLNSKLPVEILGWELSDIDHDKLD 181
Qy 75 QVEPSIAMLILKLGQYGLPSTLPPVMKQGVVAISSAFIGIASNPRLTAVAPVM 134
Db 182 ROEFAPMFLVYCALE-----KEPVPM 203
Qy 135 GSIPVGNMSPVWSSVPAVPLANGAPVIOPLPAFAHPATWPKSSFSRSGPSQI 194
Db 204 -----SLPPLVLP-----SKR 215
Qy 195 NTKLQAKQSFVYASAPPALEMAVPOSSRLKYROLFNSHDKTSGLITGQARTILMQSSL 254
Db 216 KT-----WVSPAEKAKYDEIFLTKDKMDGVYSGLEVETFLKTGL 257
Qy 255 POAQLASTINLSIDIDQDKLTAEPLTAMHLIDVAM--SQGPI--PVLYPREYIPSPFRVRS 313
Db 258 PSALLAHIMSLCDTKGCKGLSKDQFALAFHILINQKIGIDPESHITPEMIPPS---DRS 314
Qy 314 GSGMSVYSSSVQDRLPEERSSSEDEQPEKLPVTEDEKRENFERSVELEKROALLE 373
Db 315 SLQKNTGSSPV-----ADSSAIK-----ELDTLNEIYD 344
Qy 374 QORKEQERLAQLERAEQRERERQOEAKROLELEKOLEKOREROREEERKEIER 433
Db 345 LOREK-----NNVQDLKEKE-----DTVKORTSEVQ---DLQDEVORE 380
Qy 434 EAKRELEROROLEWERNROROLLNQRNKEQEGTVLAKARKLTLEFELALNDKQOLEG 493
Db 381 SINLOQLQAKOQ-----OVQELIGE-----LDQKQALE 410
Qy 494 KLQDRCRATQROEIESTNKSRELIAETIHLQOOLQESQOMLGLRLEPEKOLISQOLQ 553
Db 411 QLOEVAKKAEBAQOLISSLK-----AETSOESQISSYEBELK-----AREELR 456
Qy 554 VOONSILHRDLSLTKRALFAKELARQOLREOLDEVERE-----TRSKLOEIDVFNQLEK 608
Db 457 LOQETQO-----LEESVESGKQOLEPLQOHLQESQOEISSMQMRLEMDLETNNQ--- 507
Qy 609 LREIHSKQOLQO-----RSLFAARLKQOEQRKSLLEKQEKEDAQRVQ 653
Db 508 -----SNMSSSPQSVLVNGATDYCSLSTSSSETANFNEHAGQNNLESEPTHOESSVRSS 562
Qy 654 ER-----DKQMLEHVQOEQPRPKRPHEDRLKREDSVRKKEA-----EERAKPEMQDK 702
Db 563 PEIAPSDVTDEBAVTVAGNEKVTPR--FDDDKHSKEEDPFNVESSSLTDAVADTNLDPF 620

```

```

QY OSRL-----PHOAPAKLATQAPWSTTEKGLTISAQSVKVVY 743
Db QSDPFVSGDPFKDDPFKGKIDPFQGGDPFKSDPPASGCFKQHSIDPFSTSDPFASSN 660
QY 744 RALYPFESRSHDEITIQPGDIVMWDESQTGP--GWLGGELNGKTGMFPFANAKEIPENE 801
Db 661 SSNTSEVETWKHNDPFAFGCTVVAASDSATDPPASVFGNESFG--DGFADPFTLSKYNED 749
QY 802 VPT-----AKPTDLTSA-----PA--PKLARETPAFLPTSSSEPTPN 842
Db 740 AFNPTSSSTSVTTAKPFLLEETASGEVFPALPFPKVGTPTRPCPP-----PGRPIN 794
QY 843 WADFSST-----WPSSNEKEPETHDNDTMAAQSLTVPSAGLQKRSAPFPATATGSS 895
Db 795 KLDSSDPLKLANDPFPQFPFGNDSPEKEDPMFC-----DPFTSTTTNKE 838
QY 896 PSP 898
Db 839 ADP 841

```

RESULT 10
US-08-728-323A-2
; Sequence 2, Application US/08728323A
; Patent No. 5948676

```

1 GENERAL INFORMATION:
2 APPLICANT: Chang, Yuan
3 APPLICANT: Bohenzky, Roy A.
4 APPLICANT: Russo, James J.
5 APPLICANT: Edelman, Isidore S.
6 APPLICANT: Moore, Patrick S.
7 TITLE OF INVENTION: Immediate Early Protein From
8 TITLE OF INVENTION: Sarcoma-Associated Herpesvirus
9 TITLE OF INVENTION: Encoding Same And Uses Thereof
10 NUMBER OF SEQUENCES: 21
11 CORRESPONDENCE ADDRESSES:
12 ADDRESSEE: Cooper & Dunham LLP
13 STREET: 1185 Avenue of the Americas
14 CITY: New York
15 STATE: New York
16 COUNTRY: U.S.A.
17 ZIP: 10036
18
19 COMPUTER READABLE FORM:
20 MEDIUM TYPE: Floppy disk
21 COMPUTER: IBM PC compatible
22 OPERATING SYSTEM: PC-DOS/MS-DOS
23 SOFTWARE: PatentIn Release #1.0, Version #1.30
24 CURRENT APPLICATION DATA:
25 APPLICATION NUMBER: US/08/728,323A
26 FILING DATE:
27 CLASSIFICATION: 435
28 ATTORNEY/AGENT INFORMATION:
29 NAME: White, John P.
30 REGISTRATION NUMBER: 28, 678
31 REFERENCE/DOCKET NUMBER: 0575/52268/JPM/MSC/SKS
32 TELECOMMUNICATION INFORMATION:
33 TELEPHONE: 212-278-0400
34 TELEFAX: 212-391-0525
35 INFORMATION FOR SEQ ID NO: 2:
36 SEQUENCE CHARACTERISTICS:
37 LENGTH: 1162 amino acids
38 TYPE: amino acid
39 TOPOLOGY: linear
40
41 MOLECULE TYPE: protein
42
43 JS-08-728-323A-2

```

	Query Match	6.1%	Score 382;	DB 2;	Length 1162;
	Best Local Similarity	24.6%	Pred. No.	1.6e-18;	Gaps 19,
	Matches 127;	Conservative 139;	Mismatches 197;	Indels 54;	
QY	327 QRLRP--PPSSDEDO--PEKTLPTTFEDKKRENFERSGVLEEKRRALLFEQRKEQRL	382			
	: : : : :				

[illegible]

RESULT 11
US-09-298-568-2
; Sequence 2, Application US/09298568

```

: Patent No.65322792
: GENERAL INFORMATION:
: APPLICANT: Kieff, Elliott D.
: APPLICANT: Ballestas, Mary E.
: APPLICANT: Kaye, Kenneth M.
: TITLE OF INVENTION: RHADINO VIRUS LANA ACTS IN TRANS ON A UNIT OF RHADINO
: TITLE OF INVENTION: VIRUS DNA TO MEDiate EFFICIENT EPISOME PERSISTENCE
: FILE REFERENCE: 16412-10001R
: CURRENT APPLICATION NUMBER: US/09/298,568
: CURRENT FILING DATE: 1999-04-21
: EARLIER APPLICATION NUMBER: US 60/109,422
: EARLIER FILING DATE: 1998-11-19
: NUMBER OF SEQ ID NOS: 3
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 2
: LENGTH: 1162
: TYPE: PRT
: ORGANISM: Kaposi's sarcoma-associated herpesvirus
: US-09-298-568-2

```

```

Query Match      6.1%; Score 382; DB 4; Length 1162;
Best Local Similarity 24.6%; Pred. No. 16-18;
Matches 127; Conservative 139; Mismatches 197; Indels 54; Gaps 197

QY      327  QRLPE--EPSSDEEQ--PEKKLVPTFEDKKRNFFRGSVLEKRQALLQQRKEQRL 382
Db      551  QREPQREPPQREPPQREPPQREPPQREPPQREPPQREPPQREPPQREPPQREPPQ 608

QY      383  AQLERADEKRRERKQDEAKRQLLEKQLEKQELERQREERKKELEEREAARLEER 442
Db      609  EQ-QQDEQQQDEQQQDEQQQDEQQQDEQQQDEQQQDEQQQDEQQQDE--QQQDE 664

QY      443  QRLERERNRQELLNQRNKEDEGTIVLKKARKTLFELMLNPKQQLDEKLDIRQL 502
Db      665  QQQDEQQQDEQQQDEQQQDEQQQDE--EQQDEQQQDEQQQDEQQQDEQQQ 709

```

[illegible]

RESULT 12
US-09-410-399-2

```

: Sequence 2, Application US/09410399
: Patent No. 6482587
:
: GENERAL INFORMATION:
: APPLICANT: Robertson, Erle S.
: APPLICANT: Cotter, Murray A.
: TITLE OF INVENTION: Methods to Inhibit or Enhance the Binding of Viral DNA
: TITLE OF INVENTION: to Genomic Host DNA
: FILE REFERENCE: UM-03778
: CURRENT APPLICATION NUMBER: US/09/410,399
: CURRENT FILING DATE: 1999-10-01
: NUMBER OF SEQ ID NOS: 6
: SOFTWARE: PatentIn Ver. 2.0
:
: SEQ ID NO 2
:
: LENGTH: 1162
:
: TYPE: PRT
:
: ORGANISM: Kaposi's sarcoma-associated herpesvirus
: OS-09-410-399-2

```

[illegible]

```

Oy 679 KLKREDSVAKKAEERAKEMOD-----KOSLFIHNOBPALALQAPMSTTEKCEPLI-- 7322
Db 883 EVEEOEOQLEEEVEEOQEOQLEEEVEEOQLEEEVEEOQOQVBOQOEYEE--PIILHG 9441HG
Oy 733 -SAOESKVVYVYRALPPFSRSDEI-LIOPGDIWMVDESOTGEPWMLGELKCKTGMFP 7906
Db 942 SSSEDEMEVD-----PVVVS-THEOIASPPDONTDDDDPQPGPSREYRYVLKTSPPHRP 9959
Oy 791 ANYAEKIPENEVEYTPAKPVTDLTSAPAKALRETPA 827
Db 996 GVAMRRKVP---VTHPKKPHRYQQPPVYRQIIDDCPA 1029

```

RESULT 13
US-08-056-200-94

Sequence 94, Application US/08056200
Patent No. 5616500
GENERAL INFORMATION:
APPLICANT: Steinert, Peter M.
APPLICANT: Lee, Seung-Chul
APPLICANT: Kim, In-Gyu
APPLICANT: Chung, Soo-Il
APPLICANT: Park, Sang-Chul
TITLE OF INVENTION: Trichoyalin and Transglutaminase-3 and
METHOD OF INVENTION: Methods of Using Same
NUMBER OF SEQUENCE: 117
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe, Martens, Olson & Bear
STREET: 620 Newport Center Drive, Sixteenth Floor
CITY: Newport Beach
STATE: CA
COUNTRY: U.S.A.

```

1      ZIP: 92660
2
3      COMPUTER READABLE FORM:
4      MEDIUM TYPE: Floppy disk
5      COMPUTER: IBM PC compatible
6      OPERATING SYSTEM: PC-DOS/MS-DOS
7      SOFTWARE: PatentIn Release #1.0, Ver
8      CURRENT APPLICATION DATA:
9      APPLICATION NUMBER: US/08/056,200
10     FILING DATE: 30-APR-1993
11
12     CLASSIFICATION: 435
13
14     ATTORNEY/AGENT INFORMATION:
15     NAME: Fedrick, Michael F.
16     REGISTRATION NUMBER: 36,799
17     REFERENCE/DOCKET NUMBER: N19054.001A
18     TELECOMMUNICATION INFORMATION:
19     TELEPHONE: (714) 760-0404
20     TELEFAX: (714) 760-9502
21
22     INFORMATION FOR SEQ ID NO: 94:
23     SEQUENCE CHARACTERISTICS:
24     LENGTH: 1898 amino acids
25     TYPE: amino acid
26     TOPOLOGY: linear
27
28     MOLECULE TYPE: protein
29
30     US-08-056-200-94

```

[illegible]

Query Match 6.1%; Score 381; DB 2; Length 1898;
 Best Local Similarity 27.0%; Pred. No. 3,9e-18;
 Matches 136; Conservative 117; Mismatches 154; Indels 96; Gaps 21;

484 LNDKHOEGKLDIRCLATQROEISTNKSRELIAITHLOOQLOESQ-----OML- 537
 514 --ERREOQERREOQIR-REOERREOQLKREOERREOQLRSEQQLRREOERREOQLK 570
 538 -----GRLIPEKQILSDOL-----KOVQNSLRHDSLTLLKRALFAKELARQ 580
 571 REEKRLERREOQLKREOERREOQLKREOERREOQLKREOERREOQLKREOERREOQLK 630
 581 LREQLDE-VERETSKLOEIDVFNNQLELR---EHSKQOLQKORSL---EARKKOK 632
 631 QOEERDERLKEEPEERREHLLKSEOEERREHLOLRREOERREOQLKREOERREOQLK 690
 633 -----EOKRSLLEI-EKQKEDQORRVOERDKOMLEHVQOEBOPRPK----- 673
 691 LKREHEERREOQLRSEOEBOQERIKSRIPKQWOLESEADARQSIVLLAPQAGRAEA 750
 674 PHEEDRLKREDSVAKKEAEERAKPEMODKQSLRPHQEPKATLQAPWSTTEKGPLTIS 733
 751 POEERKRRSELSQWOEERRAHQOEBOERDFTWQWQAE-----EKSERGRQRLS 803
 QY 734 AOEIVKVVYVYALYPFESRSHDE 756
 DB 804 APPLREORERQRLAERQOQREQ 826

RESULT 14
 US-08-800-644-94
 Sequence 94, Application US/08800644
 Patent No. 5958752
 GENERAL INFORMATION:
 APPLICANT: Steiner, Peter M.
 APPLICANT: Lee, Seung-chul
 APPLICANT: Kim, In-Gyu
 APPLICANT: Chung, Soo-Il
 APPLICANT: Park, Sang-Chul
 TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and
 TITLE OF INVENTION: Methods of Using Same
 NUMBER OF SEQUENCES: 117
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Knobbe, Martens, Olson & Bear
 STREET: 620 Newport Center Drive, Sixteenth Floor
 CITY: Newport Beach
 STATE: CA
 COUNTRY: U.S.A.
 ZIP: 92660
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/800,644
 FILING DATE: 14-FEB-1997
 CLASSIFICATION: 424
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/056,200
 FILING DATE: 30-APR-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Fedick, Michael F.
 REGISTRATION NUMBER: 36,799
 REFERENCE/DOCKET NUMBER: NH054,001A
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (714) 760-0404
 TELEFAX: (714) 760-9502
 INFORMATION FOR SEQ ID NO: 94:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1898 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-800-644-94

Query Match 6.1%; Score 381; DB 2; Length 1898;
 Best Local Similarity 27.0%; Pred. No. 3,9e-18;
 Matches 136; Conservative 117; Mismatches 154; Indels 96; Gaps 21;

QY 326 DQRIPEPSSDEOQPEKLVPTFEDKRENFGSVLEKRNQALLIQ--RKEQ--R 381
 DB 348 EQQLRRE---QEERREOQLRREOERREOQLRREOERREOQLRREOERREOQLR 404
 QY 382 LAQLEAEORKEE--ROEBAKR--OLELEKOLEKRELEORQEER----- 426
 DB 405 EQQLRREOQLRREOQLRREOQLRREOQLRREOQLRREOQLRREOQLRREOQLR 464
 QY 427 ---RKEIEREBAKRELEOROLEWERNRROELINQNKQOEGTVVAKARKTLEFELA 483
 DB 465 RKRQOERRDWLKREETERH---EERRKQOL-KRDQEE-----RERWLKLEE- 513
 QY 484 LNDKHOEGKLDIRCLATQROEISTNKSRELIAITHLOOQLOESQ-----OML- 537
 DB 514 --ERREOQERREOQLR-REOERREOQLKREOERREOQLRSEQQLRREOERREOQLK 570
 QY 538 -----GRLIPEKQILSDOL-----KOVQNSLRHDSLTLLKRALFAKELARQ 580
 DB 571 REEKRLERREOQLKREOERREOQLKREOERREOQLKREOERREOQLKREOERREOQLK 630
 QY 581 LREQLDE-VERETSKLOEIDVFNNQLELR---EHSKQOLQKORSL---EARKKOK 632
 DB 631 QOEERDERLKEEPEERREHLLKSEOEERREHLOLRREOERREOQLKREOERREOQLK 690
 QY 633 -----EOKRSLLEI-EKQKEDQORRVOERDKOMLEHVQOEBOPRPK----- 673
 DB 691 LKREHEERREOQLRSEOEBOQERIKSRIPKQWOLESEADARQSIVLLAPQAGRAEA 750
 QY 674 PHEEDRLKREDSVAKKEAEERAKPEMODKQSLRPHQEPKATLQAPWSTTEKGPLTIS 733
 DB 751 POEERKRRSELSQWOEERRAHQOEBOERDFTWQWQAE-----EKSERGRQRLS 803
 QY 734 AOEIVKVVYVYALYPFESRSHDE 756
 DB 804 APPLREORERQRLAERQOQREQ 826

RESULT 15
 US-08-475-894-2
 Sequence 2, Application US/08475894
 Patent No. 5641748
 GENERAL INFORMATION:
 APPLICANT: Yen-Ming Hsu
 TITLE OF INVENTION: THE CAIP-LIKE GENE FAMILY
 NUMBER OF SEQUENCES: 6
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: LAHIVE & COCKFIELD
 STREET: 60 State Street, Suite 510
 CITY: Boston
 STATE: Massachusetts
 COUNTRY: USA
 ZIP: 02109-1875
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/475,894
 FILING DATE:
 CLASSIFICATION: 530
 ATTORNEY/AGENT INFORMATION:
 NAME: Louis Myers
 REGISTRATION NUMBER: 35,965
 REFERENCE/DOCKET NUMBER: BGP-191
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617) 227-7400
 TELEFAX: (617) 227-5941
 INFORMATION FOR SEQ ID NO: 2:

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 553 amino acids
; TYPE: amino acid
; TOPOLOGY: 1 linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: internal
US-08-475-894-2

```

```

Query Match          5.3%; Score 333; DB 1; Length 553;
Best Local Similarity 27.4%; Pred. No. 1,7e-15;
Matches 93; Conservative 67; Mismatches 115; Indels 64; Gaps 15;

```

```

QY 912 QALYPMRAKKDNLNFKSDVITVLEQD-MMMFGEVQOGKWFPSYVKLISGPYRKST 970
DB 77 QDIYPLYTRHDDDELITISVGEIITNIRKEDGXWMEGQINGRGLFPDNFVREIKEMKDP 136
QY 971 SIDTGPTESPASLKRVASPAKPAIPGEEFI-----AMTYESESQGDITF 1016
DB 137 LTNKAP-EXP--LHEV--PSGNSLSSEITLITNKRERRRRRCQVAFSYLPQNDDEL 191
QY 1017 OOGDVI-VYTKKGDWMTGTGDKSGVFPSNYVR-----LKDSE 1054
DB 192 KVGDIIEVVGVEEGWEGVLNGKTGMFSPNFIKELSGSDELGISQDEQLSKSLRETT 251
QY 1055 GSGTAGKTGSLGKPEIAQVIASVATGPEOL-----TLAPGQLILIRKKNPGGWEGE 1108
DB 252 GSESDGGDSS-STKSEGANGTATATAIQPKYKGVGFQDIFKDKPIKLRPS----IYVE 306
QY 1109 LQARGKROIG-WFPANYVKLISPGTSKITPTLPTFAVQPAVCOYIGMYDYTAQNDDEL 1167
DB 307 NDFLPVEKTIQKLPAT--TATPDSK---TEMDSRTSKDYCKVI--FPYEAQNDDEL 358
QY 1168 AFSKGOIINVLNKE--DPDMWKEVSGOGLFPSNYVKL 1204
DB 359 TIKEGDITVLTINKDCIDVGWMEGELNGRKGVPFDPNFVKL 397

```

```

Search completed: December 4, 2003, 15:21:35
Job time : 28 secs

```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 4, 2003, 15:20:14 ; Search time 34 Seconds

(without alignments)
6635.241 Million cell updates/sec

Title: US-09-674-237A-3

Perfect score: 6269

Sequence: 1 MAQFPTPGSGSLDVAITVE.....VGLFSPSNVYKLTMDPSQ 1213

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 684280 seqs, 185983659 residues

Total number of hits satisfying chosen parameters: 684280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2058.5	32.8	462	9	US-09-879-957-38
2	1388.5	22.1	530	9	US-09-879-957-194
3	1330.5	21.2	530	10	US-09-764-868-778
4	816.5	13.0	248	9	US-09-879-957-40
5	572	9.1	144	11	US-09-764-881-113
6	527	8.4	813	10	US-09-964-899-25
7	431.5	6.8	224	11	US-09-764-881-112
8	429	6.8	90	9	US-09-864-761-33620
9	429	6.8	90	9	US-09-864-761-43004
10	426	6.8	90	9	US-09-864-761-33635
11	422	6.7	82	9	US-09-864-761-46565
12	413	6.6	87	9	US-09-864-761-34169
13	409.5	6.5	168	11	US-09-764-881-128
14	382	6.1	1162	12	US-10-294-804-2
15	363.5	5.8	2701	15	US-10-171-311-83

16	343.5	5.5	843	12	US-10-144-194A-84	Sequence 84, Appl
17	333.5	5.3	779	12	US-10-144-194A-86	Sequence 86, Appl
18	330	5.3	1239	11	US-08-291-417-13	Sequence 13, Appl
19	325	5.2	659	14	US-10-144-621-2	Sequence 2, Appl
20	322	5.1	59	9	US-09-879-957-133	Sequence 133, App
21	319	5.1	735	12	US-10-144-194A-109	Sequence 109, App
22	319	5.1	784	12	US-10-144-194A-108	Sequence 108, App
23	316.5	5.0	287	10	US-09-764-868-744	Sequence 744, App
24	315.5	5.0	1170	12	US-10-341-434-95	Sequence 95, Appl
25	315.5	5.0	1564	12	US-10-144-198-2	Sequence 2, Appl
26	315.5	5.0	1564	12	US-10-144-198-4	Sequence 4, Appl
27	314	4.9	62	9	US-09-879-957-135	Sequence 135, App
28	309	4.9	56	9	US-09-879-957-134	Sequence 134, App
29	304.5	4.9	811	15	US-10-128-714-3314	Sequence 3314, Ap
30	303.5	4.8	611	9	US-09-216-393-81	Sequence 81, Appl
31	303.5	4.8	611	12	US-10-321-856-81	Sequence 81, Appl
32	300.5	4.8	1175	10	US-09-771-161A-224	Sequence 224, App
33	300.5	4.8	1175	10	US-09-771-161A-225	Sequence 225, App
34	300.5	4.8	1175	10	US-09-771-161A-226	Sequence 226, App
35	300.5	4.8	1212	12	US-10-247-671-157	Sequence 157, App
36	299	4.8	1233	11	US-08-291-417-89	Sequence 89, Appl
37	297	4.7	1070	15	US-10-128-714-8314	Sequence 8314, Ap
38	296	4.7	57	9	US-09-879-957-136	Sequence 136, App
39	295.5	4.7	1135	15	US-10-309-851-12	Sequence 12, Appl
40	295	4.7	1200	12	US-10-094-749-2682	Sequence 2682, Ap
41	294.5	4.7	2383	15	US-10-082-830-260	Sequence 260, App
42	294	4.7	1133	15	US-10-309-851-14	Sequence 14, Appl
43	294	4.7	1133	15	US-10-309-851-16	Sequence 16, Appl
44	293	4.7	888	11	US-09-893-519A-73	Sequence 73, Appl
45	293	4.7	929	11	US-09-298-523B-60	Sequence 60, Appl

ALIGNMENTS

RESULT 1
US-09-879-957-38
Sequence 38, Application US/09879957
Patent No. US20020034755A1
GENERAL INFORMATION:
APPLICANT: SPARKS, Andrew B.
HOFFMAN, No. US20020034755A1h
KAY, Brian K.
FOMKES, Dana M.
MCCONNELL, Stephen J.
TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND USING SAME
NUMBER OF SEQUENCES: 227
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/879, 957
FILING DATE: 13-Jun-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/630, 915
FILING DATE: 03-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Mastrock, S. Leslie
REGISTRATION NUMBER: 18, 872
REFERENCE/DOCKET NUMBER: 1101-174
TELECOMMUNICATION INFORMATION:

```

? TELEPHONE: (212) 790-9090
? TELEFAX: (212) 869-8864/9741
? TELEX: 66141 PENNIE
? INFORMATION FOR SEQ ID NO: 38:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 462 amino acids
? TYPE: amino acid
? STRANDEDNESS: <Unknown>
? TOPOLOGY: unknown
? MOLECULE TYPE: peptide
? SEQUENCE DESCRIPTION: SEQ ID NO: 38:
US-09-879-957-38

```

Query Match	32.8%;	Score 2058.5;	DB 9;	Length 462;
Best Local Similarity	79.3%;	Pred. No. 2.6e-101;		
Matches 391;	Conservative 13;	Mismatches 18;	Indels 71;	Gaps 1;

OY		721	W S T E R K P L I T A O E S K V V Y R A L Y P E S H E I T O P D I Y M V E S Q G E G M G	780
Dd		41	F O L O K P L I T A O E S K V V Y R A L Y P E S H E I T O P D I Y M V E S Q G E G M G	100
OY		781	E L K G T G F P N A Y A E I P E N E V T P A K P V T D I T S A P A K A L R E T P A L V T S S P T P	840
Dd		101	E L K G T G F P N A Y A E I P E N E V P A V K E V T D I T S A P A L R E T P A L V T S S P T P	160
OY		841	N N A D F S T W P S S N E K P E T D W D P A O P S I T V S A G O L R O R S A F T P A T A N G S P S V L	900
Dd		161	N N A D F S T W P S S N E K P E T D W D P A O P S I T V S A G O L R O R S A F T P A T A N G S P S V L	220
OY		901	G O G E V E G I Q A O A L Y P W R A K D N H N F N K S D V I T V L E O D W W F G E V O G O K M P K S Y Y K	960
Dd		221	G O G E V E G I Q A O A L Y P W R A K D N H N F N K S D V I T V L E O D W W F G E V O G O K M P K S Y Y K	280
OY		961	L I S G V R K S T I D N C P T S P A S I K R V A S P A A P A L R G E F I A N Y E S B O G D L F Q G D	102
Dd		281	L I S G I R K S T I D N C P T S P A S I K R V A S P A A P A L R G E F I A N Y E S B O G D L F Q G D	318
OY		1021	V I V V T K D G D W T G V G D K G V P S N Y R L K D S B G S G A G T G S I G K K E I A O V I A S Y A A	1080
Dd		319	-----E I A O V I A S Y A	329
OY		1081	T G P O U T A P G O L I I R K N P G W E G I O A R G K R O I G P P A N Y V K L S P C S I T P E	1144
Dd		330	T G P O U T A P G O L I I R K N P G W E G I O A R G K R O I G P P A N Y V K L S P C S I T P E	389
OY		1141	L P K T A V O P A V C V I G M Y D T A O N D E L A F S K O I I N V L K E D P D W K G E V S Q V E L P S N	1200
Dd		390	P P K T A V O P A V C V I G M Y D T A O N D E L A F N K Q I I N V L K E D P D W K G E V S Q V E L P S N	449
OY		1201	Y V K L T T D D P S Q	1213
Dd		450	Y V K L T T D D P S Q	462

US-09-879-957-194
Sequence 194, Application US/09879957
Patent No. US20020034755A1
GENERAL INFORMATION:
APPLICANT: SPARKS, Andrew B.
HOFFMAN, NO. US20020034755A1h
KAY, Brian K.
FOWLKES, Dana M.
McCONNELL, Stephen J.
TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
USING SAME
NUMBER OF SEQUENCES: 227
CORRESPONDENCE ADDRESS:
ADDRESSEE: Femite & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York

1 COUNTRY: USA
2 ZIP: 10036-2711
3
4 COMPUTER READABLE FORM:
5
6 MEDIUM TYPE: Floppy disk
7
8 COMPUTER: IBM PC compatible
9 OPERATING SYSTEM: PC-DOS/MS-DOS
10 SOFTWARE: PatentIn Release #1.0, Version #1.30
11
12 CURRENT APPLICATION DATA:
13 APPLICATION NUMBER: US/09/879,957
14 FILING DATE: 13-Jun-2001
15
16 CLASSIFICATION: <Unknown>
17
18 PRIOR APPLICATION DATA:
19 APPLICATION NUMBER: US 08/630,915
20 FILING DATE: 03-APR-1996
21
22 ATTORNEY/AGENT INFORMATION:
23 NAME: Mistrock, S. Leslie
24
25 REGISTRATION NUMBER: 18,872
26
27 REFERENCE/DOCKET NUMBER: 1101-174
28
29 TELECOMMUNICATION INFORMATION:
30 TELEPHONE: (212) 790-9090
31 TELEFAX: (212) 869-8864/9741
32
33 TELEX: 66141 PENNIE
34
35 INFORMATION FOR SEQ ID NO: 194:
36
37 SEQUENCE CHARACTERISTICS:
38
39 LENGTH: 509 amino acids
40
41 TYPE: amino acid
42
43 STRANDEDNESS: <Unknown>
44
45 TOPOLOGY: unknown
46
47 MOLECULE TYPE: peptide
48
49 SEQUENCE DESCRIPTION: SEQ ID NO: 194:
50
51 US-09-879-957-194

Query	Match	Similarity	22.1%	Score	1388.5	DB	9	Length	509
Best Local	Similarity	48.9%	Pred. No.	8.3e-66					
Matches	287	Conservative	85	Mismatches	124	Indels	91	Gaps	14
QY	635	ERKSLTELEKOR	----	EDAQRRVQERDQOMLEHVOQEBQPRPRKPHBEDRLKREDSVRKKE	690				
Db	6	ERKSLTELEKOR	----	EDAQRRVQERDQOMLEHVOQEBQPRPRKPHBEDRLKREDSVRKKE	690				
QY	691	AEEPAKPEPMODKOSRLFHHPHOEPKALQOAWSTTEKGPRLTISAQSVKVVYRYALPYPE	750						
Db	60	AEEPAKPEPMODKOSRLFHHPHOEPKALQOAWSTTEKGPRLTISAQSVKVVYRYALPYPE	750						
QY	751	SRSDEITLQPGDIVMVDQSGTQGEPMGLKGTGMPFANYAKIP--ENEVTPPAKP	808						
Db	87	ARNDEHNSFNSGDIIOVDEKTVGEBGMILYSGFOGNGFMPFCNYEKMSSENE-----	139						
QY	809	VTDLTSAAPKALRETAPLPVTSSEBITPNNMADESSWTSPSSNEKEPFTMDMTAA	868						
Db	140	VTDLTSAAPKALRETAPLPVTSSEBITPNNMADESSWTSPSSNEKEPFTMDMTAA	868						
QY	869	QPSLTIVSAGLQROBSAFTPATATGSSSPVLGCGEKEGLOALYWRAKKNDHNLFPN	928						
Db	181	FSNLTIVTWSQ--KKSAPTRTVSPG-SVSPHGGQGVVENLKAQALCMTAKDHLNFS	237						
QY	929	KSDVITLLEQDDMMFGEVQOKGMPFSYKYLISGPRKSTSIDTGPTESPASLRVAVS	988						
Db	238	KHDITVLLEQDDMMFGEVQOKGMPFSYKYLISGPRKSTSIDTGPTESPASLRVAVS	988						
QY	989	--PAKPAIPGEELIANTYTESSEBQDITFOQGDVIVTKKDGDMWTGVDGKSGVPPSN	1046						
Db	290	KKPTSAASVGEELIANTYTESSEBQDITFOQGDVIVTKKDGDMWTGVDGKSGVPPSN	1046						
QY	1047	YVRLKDSBGSITAGKTGSLGKKPELQVITASYAATGPPQILILAPQILILIKKNTGGMWE	1100						
Db	350	YVRLKDSBGSITAGKTGSLGKKPELQVITASYAATGPPQILILAPQILILIKKNTGGMWE	1100						
QY	1107	GELDARGKKRQIGMFANYVTLSPGTSTKIPTELEPKTAVGPAYGCVIGMTDYTQONDE	1166						
Db	410	GELDARGKKRQIGMFANYVTLSPGTSTKIPTELEPKTAVGPAYGCVIGMTDYTQONDE	1166						
QY	1167	LAFSGKQIINVLNKEDPDMMKGEVSGQVGLFPPSNVYKLTITMDPSQQ	1213						


```

; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (183)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (202)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-881-132

Query Match
Best Local Similarity 45.6%; Score 431.5; DB 11; Length 224;
Matches 103; Conservative 31; Mismatches 51; Indels 41; Gaps 8;

QY 115 FGIGGASMPPLTAVAPVPMG-----STPVGMSPVLSVPPAAVPLANG 161
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 9 FPMG---SMPNLSTPQPLPPAPATSLSATSGTNLPMLMPTPLVPSVSTSLP---NG 62

QY 162 APPVIOPLPAPAPATWPKSSFS--RSGPGSQLMTKLQAKS--PDVASAP----- 211
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 63 TMSLQPLP-IPYSSSTLPHGSSYSILMGFG---GASIQAKSLIDLSSSSSTSTASL 118

QY 212 -----AAEVAVPOSSSLKTRQLFNSHDKTWSGHLTPQARTILMOSSLPPQAOIASIW 263
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 119 SGNSEPTGSEWAVPQTRLKTRQKENTLIDKSMGVLGFGQARNALLQSNLSQTOLATIW 178

QY 264 NUSDIDDOCKLTAEFTILMHLIDVAMSGQPLPVUPPEVYIPPSFR 309
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 179 TLADXDGDQLKAEFTILMHLTMKAKQPLPLTLPELVPSFR 224

RESULT 8
US-09-864-761-33620
; Sequence 33620, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David K.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecm1ca-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263,6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661

```

```

; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 33620
; LENGTH: 90
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AP000193.1
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.3
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.2
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.7
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.2
; OTHER INFORMATION: EST_HUMAN HIT: A4773823.1, EVALUE 2.00e-37
; OTHER INFORMATION: SWISSPROT HIT: Q15811, EVALUE 2.00e-45
US-09-864-761-33620

Query Match
Best Local Similarity 94.0%; Score 429; DB 9; Length 90;
Matches 79; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 767 VDESQGERGWLGLKGTGWFPPANYAEKIPENVEPTPAKPVTLTSAPAKLARETP 826
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 1 VDESQGERGWLGLKGTGWFPPANYAEKIPENVEPTPAKPVTLTSAPAKLARETP 60

QY 827 APLVTSSEPTTPNNMADFSSSTW 850
||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 61 APLVTSSEPTTPNNMADFSSSTY 84

RESULT 9
US-09-864-761-43004
; Sequence 43004, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecm1ca-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263,6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664

```

```

PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00669
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00665
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00668
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00663
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00662
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00661
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00670
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: US 60/234,687
PRIORITY FILING DATE: 2000-09-21
PRIORITY APPLICATION NUMBER: US 09/608,408
PRIORITY FILING DATE: 2000-06-30
PRIORITY APPLICATION NUMBER: US 09/774,203
PRIORITY FILING DATE: 2001-01-29
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 43004
LENGTH: 90
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AP000311.1
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 0.87
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.3
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 32
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.66
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.92
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.98
OTHER INFORMATION: EST_HUMAN HIT: AAT73823.1, EVALUATE 2.00e-37
OTHER INFORMATION: SWISSPROT HIT: Q15811, EVALUATE 2.00e-45
US-09-864-761-43004

Query Match
Best Local Similarity 6.8%; Score 429; DB 9; Length 90;
Matches 79; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Cy 767 VDESGTGPGLGELKGTGFWPANYAEKIPENVEPTPAKVDLTSAAPKALRETP 826
Db 1 VDESGTGPGLGELKGTGFWPANYAEKIPENVEPTPAKVDLTSAAPKALRETP 60

Cy 827 APLPVTSEPTSTPNMADFSSTW 850
Db 61 APLAVTSEPTSTPNMADFSSTW 84

RESULT 10
US-09-864-761-33635
Sequence 33635, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecm1ca-X-1
CURRENT FILING DATE: 2001-05-23
CURRENT APPLICATION NUMBER: US/09/864,761
PRIORITY FILING DATE: 2000-02-04
PRIORITY APPLICATION NUMBER: US 60/180,312
PRIORITY FILING DATE: 2000-05-26
PRIORITY APPLICATION NUMBER: US 60/207,456
PRIORITY FILING DATE: 2000-05-26
```

```

PRIORITY APPLICATION NUMBER: US 09/632,366
PRIORITY FILING DATE: 2000-08-03
PRIORITY APPLICATION NUMBER: GB 24263.6
PRIORITY FILING DATE: 2000-10-04
PRIORITY APPLICATION NUMBER: US 60/236,359
PRIORITY FILING DATE: 2000-09-27
PRIORITY APPLICATION NUMBER: PCT/US01/00666
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00667
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00664
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00669
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00665
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00668
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00663
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00662
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00661
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00670
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: US 60/234,687
PRIORITY FILING DATE: 2000-09-21
PRIORITY APPLICATION NUMBER: US 09/608,408
PRIORITY FILING DATE: 2000-06-30
PRIORITY APPLICATION NUMBER: US 09/774,203
PRIORITY FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 33635
LENGTH: 99
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AP000311.1
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.2
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 3.2
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.6
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.3
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.2
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.4
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.7
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.2
OTHER INFORMATION: SWISSPROT HIT: O35601, EVALUATE 2.80e-01
OTHER INFORMATION: EST_HUMAN HIT: AAB15076.1, EVALUATE 1.00e-08
US-09-864-761-33635

Query Match
Best Local Similarity 6.8%; Score 426; DB 9; Length 99;
Matches 86; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Cy 348 TFEDEKKEENFEGSVLEKRRQALIEQKREKRLAERAEQKREKREKREKREKREK 407
Db 1 TFEDEKKEENFEGSVLEKRRQALIEQKREKRLAERAEQKREKREKREKREKREK 60

Cy 408 LEKOLEKORELEKREKREKREKREKREKREKREKREKREKREKREKREKREKREK 439
Db 61 LEKOLEKORELEKREKREKREKREKREKREKREKREKREKREKREKREKREKREK 92

RESULT 11
US-09-864-761-46565
Sequence 46565, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
```

APPLICANT: Rank, David R.
 APPLICANT: Hanzel, David K.
 APPLICANT: Chen, Wensheng
 TITLE OF INVENTION: HUMAN GENOME- DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
 TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
 FILE REFERENCE: Aemica-X-1
 CURRENT APPLICATION NUMBER: US/09/864,761
 CURRENT FILING DATE: 2001-05-23
 PRIOR APPLICATION NUMBER: US 60/180,312
 PRIOR FILING DATE: 2000-02-04
 PRIOR APPLICATION NUMBER: US 60/207,456
 PRIOR FILING DATE: 2000-05-26
 PRIOR APPLICATION NUMBER: US 09/632,366
 PRIOR FILING DATE: 2000-08-03
 PRIOR APPLICATION NUMBER: GB 24263.6
 PRIOR FILING DATE: 2000-10-04
 PRIOR APPLICATION NUMBER: US 60/236,359
 PRIOR FILING DATE: 2000-09-27
 PRIOR APPLICATION NUMBER: PCT/US01/00666
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00667
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00664
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00669
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00665
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00668
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00663
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00662
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00661
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00670
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: US 60/234,687
 PRIOR FILING DATE: 2000-09-21
 PRIOR APPLICATION NUMBER: US 09/608,408
 PRIOR FILING DATE: 2000-06-30
 PRIOR APPLICATION NUMBER: US 09/774,203
 PRIOR FILING DATE: 2001-01-29
 NUMBER OF SEQ ID NOS: 49117
 SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
 SEQ ID NO 46565
 LENGTH: 82
 TYPE: PRT
 ORGANISM: Homo sapiens
 FEATURE:
 OTHER INFORMATION: MAP TO AP000117.1
 OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.7
 OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.4
 OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.2
 OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 0.76
 OTHER INFORMATION: EXPRESSED IN HEL100, SIGNAL = 1.3
 OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.5
 OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1
 OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.95
 OTHER INFORMATION: SWISSPROT HIT: Q15811, EVALUATE 9.00e-45
 OTHER INFORMATION: EST_HUMAN HIT: AA773823.1, EVALUATE 2.00e-37
 US-09-864-761-46565
 Query Match 6.7%; Score 422; DB 9; Length 82;
 Best Local Similarity 95.1%;
 Matches 78; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 767 VDESOTGEPWIGSELKSGKTGWFPANVAEKIPENEVPPTPAKPVDTLSAPAKLARETP 826
 |||||
 Db 1 VDESQTGEPWIGSELKSGKTGWFPANVAEKIPENEVPAPVVKPVDTSTSPAPKALARETP 60
 |||||
 QY 827 APLPVTSSEPTTPNNMADFSS 848

```

Db          61  APLAVTSSEPTTNNMADFSS 82

RESULT 12
US-09-864-761-34169
Sequence 34169, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecmica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 34169
LENGTH: 87
TYPE: PRT
ORGANISM: Homo sapiens
FEATURES:
OTHER INFORMATION: MAP TO AP00049.1
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.3
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.2
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.3
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 3
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.9
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.2
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.5
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 5.3
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.9
OTHER INFORMATION: EST HUMAN HIT: AA815076.1, EVALUE 4.00e-09

```


OTHER INFORMATION: SWISSPROT HIT: O35601, EVALU 1.80e-01
US-09-864-761-34169

Query Match 6.6%; Score 413; DB 9; Length 87;
Best Local Similarity 96.6%; Pred. No. 4.5e-15;
Matches 84; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 348 TPEKKRENERGVELEKROALLFOORKEOEELAEAEOREREROEAKROLE 407
DB 1 TPEKKRENERGVELEKROALLFOORKEOEELAEAEOREREROEAKROLE 60

QY 408 LEKOLEKORELEEREREERERERRE 434
DB 61 LEKOLEKORELEEREREERERERRE 87

RESULT 13
US-09-764-881-128
Sequence 128, Application US/09764881
Publication No. US20030125246A9
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PT207
CURRENT APPLICATION NUMBER: US/09/764,881
CURRENT FILING DATE: 2001-01-17
Prior application data removed - refer to PALM or file wrapper
NUMBER OF SEQ ID NOS: 192
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 128
LENGTH: 168
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (122)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (76)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (155)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (147)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (147)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (159)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (165)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-881-128

Query Match 6.5%; Score 409.5; DB 11; Length 168;
Best Local Similarity 60.4%; Pred. No. 1.5e-14;
Matches 87; Conservative 18; Mismatches 32; Indels 7; Gaps 4;

QY 1 MAOPPTFGSLDVAIVVERAKDQOFLSLPAGITTDQANFFQSGLRQPIVLAQ 60
DB 8 MAOPPTANNGGPNXKAITSEERTKGRQFNLKPSGGVITDQANFXLQSGLRPAPVLA 67

QY 61 IVALADNMNDGRMPDQVEFSIAMKLIKLIKLGQYQLPSTLPVWKQOPV--AISAAPAFGIG 118
DB 68 IVALSDLNXXKXGKMOQSFISAMKLIKLIKLGQYQLPVTLPIMKQPPMSPPLISA-RFIMG 126

QY 119 GIASMPPLTAVAPVPMGSIPIVGM 142
DB 127 ---SMPNLSIPQPLPAA-PITSL 146

RESULT 14
US-10-294-804-2
Sequence 2, Application US/10294804
Publication No. US2003013948A1
GENERAL INFORMATION:
APPLICANT: Robertson, Erle S.
TITLE OF INVENTION: Methods to Inhibit or Enhance the Binding of Viral DNA
FILE REFERENCE: UM-03778
CURRENT APPLICATION NUMBER: US/10/294,804
CURRENT FILING DATE: 2002-11-14
PRIOR APPLICATION NUMBER: US/09/410,399
PRIOR FILING DATE: 1999-10-01
NUMBER OF SEQ ID NOS: 6
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2
LENGTH: 1162
TYPE: PRT
ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-10-294-804-2

Query Match 6.1%; Score 382; DB 12; Length 1162;
Best Local Similarity 24.6%; Pred. No. 4e-12;
Matches 127; Conservative 139; Mismatches 197; Indels 54; Gaps 19;

QY 327 QRLPE--EPSSSEDOQ--PEKULPVTPEKRENERGVELEKROALLFOORKEOERL 382
DB 551 QREFOQREPOQREPOQREPOQREPOQREPOQREPOQREPOQREPOQREPOQ 608

QY 383 AOLERAEOERKEEREOEAKROLELEKOLEKORELEEREREERERERERAKRELER 442
DB 609 EQ-QODEQODEQODEQODEQODEQODEQODEQODEQODEQODEQODEQODEQ 664

QY 443 QROLEWRNRROELINQNRKEQEGTVVLKARRTLFELEALNDKQKQLEGLQODIRCL 502
DB 665 QQQDEQODEQODEQODEQODEQODEQODEQODEQODEQODEQODEQODEQODE 709

QY 503 ATQROEISTENKRELEIAITHLQOOLQESQMLGLILEKQILSLQKQVQNSLHRD 562
DB 710 DEBOEQODEQODEQODEQODEQODEQODEQODEQODEQODEQODEQODEQODE 767

QY 563 SLTLKRALAKELARQLEOLDEVERETRSKLOEIDVNNOLKELREIHSKQOLQOQ 621
DB 768 ELEBOELEDQOELEBOELEBOELEBOELEBOELEBOELEBOELEBOELEBOE 822

QY 622 RSLR--AARLKOEERKSLLEKOKEDAQRVQOERDKQWLEHYQOEOPRPRKPHR 678
DB 823 QLEBOEQELEBOEQELEBOEQELEBOEQELEBOEQELEBOEQELEBOEQELE 882

QY 679 RLKREDSVRKKEAEERAKPEMOD---KQRLRPHQOEPKATQAPWSTTEKPLTI-- 732
DB 883 EVEBOEQELEBOEQELEBOEQELEBOEQELEBOEQELEBOEQELEBOEQELE 941

QY 733 -SAQESVYVYVYALYFESRSHDEI-TIOPGDIWVDESOTGEPGLGELKGTGMP 790
DB 942 SSSSEDEVD---YVVS--THEQIASPPGDDTPDDQPGPSRERYVLRKTSPPHR 995

QY 791 ANYAEKIPENEVPTPAKPVDTLSAPAPKLALRETPA 827
DB 996 GVRMRKVP---VTHPKKPHRYQOPVPIYRQIDDCPA 1029

RESULT 15
US-10-171-311-83
Sequence 83, Application US/10171311
Publication No. US20030087270A1
GENERAL INFORMATION:
APPLICANT: Schlegel, Robert
APPLICANT: Chen, Yan
APPLICANT: Zhao, Xumei
APPLICANT: Monahan, John

```

; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Glatc, Karen
; APPLICANT: Ganavavaru, Manjula
; APPLICANT: Hoersch, Sebastian
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; TITLE OF INVENTION: OF CERVICAL CANCER
; FILE REFERENCE: MRI-035
; CURRENT APPLICATION NUMBER: US/10/171,311
; CURRENT FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: US 60/298,159
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,155
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/335,936
; PRIOR FILING DATE: 2001-11-14
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 83
; LENGTH: 2701
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-171-311-83

Query Match      5.8%; Score 363.5; DB 15; Length 2701;
Best Local Similarity 20.7%; Pred. No. 1e-10;
Matches 280; Conservative 165; Mismatches 490; Indels 415; Gaps 59;

QY 20 ERBAKHDOF--LSLKP--IAGFTIGDQARNFFFGSLPQVLAQIMALADNNNDGRMD 74
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 155 KKEKTDNDYGRPSLRPPVACWRDGRAGASPSDDDEKLPGQ-----BESTAGTSE 209
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 75 Q-----VEFSIAMKLKIKLOGYLP-----STLPVWKOQPAVISAAPAFIGIGIAS 122
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 210 QNDILKVKRKRIACGPPQAKLNGQALASQIRAMMPPTMFOQYPRMTYPPPHG----- 263
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 123 MBPLTAVALPVGMSLIPVGMSPPLVSSVPPAAPPPLANGAPVYIOPLPAPAPATWPKS 182
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 264 -----PM-----RPPSLSETKKG-----LRGRGP-----PSM-----ASEPFR 293
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 183 SSFSSRSGPSQANTKLOKQASPDVASAPPAEMAVPOSSRLKRYOLFNSHDKTMSGHLTG 242
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 294 PSILSA-----SELKEIDKFNDLDAEADGWAQAQ-MEVDYEQINFSDDEQGS-NS 344
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 243 PQARTILMOSLPQAQALASIMWLSIDIDQOKLTAEEFILAMHLIDVAMSGQPLPVLPE 302
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 345 PKENNSDDGS-----KASENNENKKTETDE-----VSNTKSSSQITP----- 380
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 303 YIPPSFRVRSVSGMSVSISSSVQRLPREP-----S 334
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 381 -AOPSVAKYPIYKGPSPFNQERGTSSHLPPPKLLAQOHPPPDROAVPGRPGPPSKQVA 439
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 335 SDE-----QQPEKLPVTEDEKKNFERSGVELERQALLERQORKEORLAQLER 387
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 440 DDEDEIKQRRROQSEISAVERARKREBEER-RME-EQRKAACAKLRLDEKLGITLX 497
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 388 --AEQKREKREQOQAKQOLEK-----QLEKQELERQREBERKKEITEREAAKRELE 441
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 498 OPSPEIRERER-EKREKREKLEKQEQEREREKREKROQEKKELEKQEKOREME 556
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 442 RQR-----QLEWERNRROELINORKEQESTVVLKARKTLEFELALNDKQHLQLEGLQD 497
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 557 KERKQEKELERQEKELQKMEQEKELQKMEQEKELERKLE---EKIEPREPRLBPMVEK 613
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 498 INCRLATOROE-----TESTNKSREILIAETIH----- 525
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 614 QSENSCNKEEBEVFTRODSNRS-EKEATPVVHETEPESSQOPRPAVLGTFQFOKSLP 672
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 526 -----LQQQLQESQO----- 535
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 673 PRFORQOQOQOQOQOQOQOQVLPQTVPSQPSSTVPPPHRPLVQPMQPHQHLASMG 732
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 536 -----MLGR-----LIPKQILSDQLKQVQNSLH----- 560
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

DB 733 FDBRMLMQSYMDPRMMSGRPAMDIPRIHGMILPPKPLMRROQMGSPSSSEFHIARS 792
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 561 -RDSLLITLRA-----LEAKEIARQOLREQLEVERETSKLOET---DVFN 604
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 793 AADHAISISEPRLMGSDPYPHAEQOATTPRATSEPEPDRSEALDQOITAAVSVEHN 852
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 605 OKE-----LREHISKQOLQK--QRSLEAALTKQKQERKSLLEKQKEDQORVQOED 656
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 853 QLEAPKADFIRE-SSEAQVQKFLRSVEDVAPPHITDANNQACFADPQKTLAPQF-- 909
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 657 KOMLEHVQOEQPRPRK--PHEEDRLKREDSVR-----KXEAERAPX 697
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 910 -----ERISAVESQPRKRSVSHGSHHTQKPDQREBPASGIRPKYTSRCLDSKEPIER--P 963
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 698 EMQDKQSLFHPHQBPALQAPMSTTEKGLTISAQESVAVVYRALYPRESSHDEI 757
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 964 EKKPKKEGFIRSESGPKPEKVKSKSETWGRP-----SSNRREBV 1005
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 758 TIQP-----GDI-----VMVDESQTEGPGWLQGLKGTQGMFPANYAEKIPENEVPTPAK-FV 809
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1006 NDRPVRRSGLPKKPVLRDKKEER-----QKREKQ-----EKAERKVEKVVAKPEKTEK 1055
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 810 TDLTSAPAPKALARETPAPLPVTSSEPSPTPNNMADFSSSTWPSSENEKETDNMDTMAAQ 869
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1056 KDLPPPPPPP-----QPPAPIQPSVPPPIQPEA-EKPFETETATLAQKRSQPT--EKPL 1108
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 870 PSLTV--PSAGQLRQSAFTPATATGSSPSVLCQGEKVEGLQAOALYPMRAKDNHIN 926
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1109 PVSTVQVEPAVTVNQITAAAVVKEKQPEKVISKDLVIE-----RPRDSRPA 1158
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 927 FKKSDVITLBEQDDMMNGEVOGQKWP-----KSYKLISGVPKST 970
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1159 VKKE--STLPPRTYW---KEAREKDFPDQYRGGRGGEYYSRGRSYGSGYGRGG- 1211
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 971 SIDGPTESPASLKEVASPAKPAIPGEFFIAMTYESESQGLDTFOQGDVIVTVTKQGD 1030
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1212 ---RGHTND-----YPOYDNKPRAEHIPSGLRQRESEETSESSEDFEVVRKRRQ 1260
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1031 WMTGVGDSGVPSNNYVRLKDSBGSAGTGSGLKPKPEIAQVIASTAAATGPQULTAP 1090
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1261 RGSETDITD-----SEIHESASDKDSLKSKLPKREERPENKPKVPHSSFKPDNHVRID 1314
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1091 GOLL---LIR---KKNPGWMEGELQARGK 1114
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1315 NRLLEKPYVRDDDKAKPGFLPKGEPTRRGR 1344
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Search completed: December 4, 2003, 15:25:42
Job time : 36 secs

THIS PAGE BLANK (USPTO)